Inferences in Semi-parametric Fixed and Mixed Models for Longitudinal Discrete Data

by

© **Nan Zheng**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics and Statistics

Memorial University

September 2017

St. John's, Newfoundland and Labrador, Canada

Abstract

Longitudinal data analysis for discrete such as count and binary data has been an important research topic over the last three decades. With regard to inferences for this type of data, the marginal model approach using 'working' correlation based GEE (generalized estimating equation), and an auto-correlation class based GQL (generalized quasi-likelihood) approach have been used, among others. This later GQL approach was suggested because of certain efficiency drawbacks of the GEE approach. Many studies were also done using the GQL approach for longitudinal mixed models. In this thesis, we study the longitudinal count and binary data in a wider semi-parametric longitudinal fixed and mixed model setup. For inferences, the SQL (semi-parametric quasi-likelihood), SGQL (semi-parametric generalized quasi-likelihood) and SML (semi-parametric maximum likelihood) have been used wherever appropriate. The asymptotic properties such as consistency of the estimators produced by these approaches have been studied in detail. We also study the finite sample properties of the new approaches and compare them where applicable with existing SGEE (semi-parametric generalized estimating equation) approaches. The proposed models and the estimation methodologies are also illustrated with some real life data.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor professor Brajendra Sutradhar, for his guidance to complete this thesis. I am thankful to him for motivating me to work in this semi-parametric longitudinal data analysis area. His support, vast knowledge and logical way of thinking have been invaluable throughout my work.

I would also like to thank professor J Concepción Loredo-Osti, my supervising committee member and appointed supervisor after the retirement of professor Sutradhar, for his advices, encouragement and financial support during my PhD project.

Thanks to all my friends and some of my fellow students for their encouragement and moral support during my program.

A special thanks goes to the Department of Mathematics and Statistics at Memorial University of Newfoundland for providing financial and other necessary opportunities, especially all administrative staff members for the help they offered me during my study.

I am so indebted to my parents and my brother. Without my family's understanding and love, I would not have been able to finish this thesis.

Table of contents

List of tables

List of figures

Chapter 1

Background of the problem

Generalized linear models (GLMs) both in independent and longitudinal contexts have been an important research topic over the last three decades. The main purpose of these types of models is to examine the effects of certain fixed covariates on the responses in either an independent or longitudinal framework. For example, in a GLM setup for longitudinal data, repeated binary data consisting of asthma status (0 or 1) collected from 537 children over a period of four years has been analyzed by many authors (Zeger et al., 1988, Sutradhar, 2003). For this problem, the main objective is to find the effect of mothers' smoking habit (a fixed covariate) on the asthma status of the children while taking the longitudinal correlation of the responses into account. For longitudinal count data a similar GLM has been fitted to various data sets by some authors. For example, we refer to the Health Care Utilization (HCU) data (Sutradhar, 2003) where repeated numbers of yearly physician visits were studied as a function of various covariates such as gender, education level, chronic disease status, and age of the individuals.

In many cases, the repeated responses can be influenced by a latent individual random effect, then their means, variances and correlation structure will also be affected by the distribution of the random effect. So, for a better modeling of the effects of the fixed covariates, it is necessary in such situations to extend the GLM to the GLMM (generalized linear mixed model) (Sutradhar, 2010) by introducing an individual random effect into the models. For example, Sutradhar and Bari (2007) revisited the HCU data by fitting it to a GLMM for longitudinal data, and obtained a better estimation of the mean and variance as compared to that from fitting a GLM for longitudinal data. For binary data, for example, Sutradhar et al. (2008) fitted a mixed model for longitudinal data to the SLID (Survey of Labour and Income Dynamics) data collected by Statistics Cananda from 1993 to 1998, for evaluating the effects of the covariates including gender, age, geographic location, education level, and marital status on the employment status (1 for 'unemployed all year', 0 for otherwise) of 15,731 individuals over a period of four years from 1993 to 1996.

For clarity, we provide these models , i.e. GLM and GLMM under an independent setup, and GLM and GLMM under a longitudinal setup in Sections 1.1 and 1.3, respectively.

The aforementioned models, GLM and GLMM in both independent and longitudinal setup, may fall short in situations where the fixed covariates used in these models may not be able to adequately explain the responses. To tackle this situation, there are studies in the literature where a secondary covariate that may not be of direct interest but may influence the responses is introduced. For example, in a longitudinal respiratory infection (binary) status study (Lin and Carroll, 2001, Section 8) gender and vitamin A deficiency status were considered as primary covariates, whereas the age effect of an individual was not of direct interest but it was included as a secondary covariate. In count data setups, one may again refer to the HCU data where similar to the respiratory infection status study, the age covariate could also be considered as a secondary covariate. In general, the effects of such secondary covariates are nonparametrically taken care of, and the GLMs in both independent and longitudinal setup are extended to SGLMs (semi-parametric generalized linear models) in both independent and longitudinal setup respectively. In Sections 1.2 and 1.4, we will provide a brief introduction of SGLMs in independent and longitudinal setup, respectively. The SGLMs for longitudinal data have been studied by some authors such as Severini and Staniswalis (1994), Lin and Carroll (2001, 2006), Warriyar and Sutradhar (2014), Sutradhar et al. (2016). However, there are some issues with the inference techniques used by some of the above authors. Also, in practice, it may happen that, in addition to the primary and secondary covariates used to construct the above mentioned SGLM for longitudinal data, the repeated responses of an individual may also be influenced by another individual latent effect. However, the analysis of this type of longitudinal responses affected by both random effects and nonparametric functions is however not adequately addressed in the literature. This thesis is aimed to address these issues. To be specific, the objective of the thesis are: (1) Inferences in the SGLM setup for longitudinal data, (2) extension of the SGLM for longitudinal data to the SGLMM setup, and (3) development of inferences under the SGLMM setup for longitudinal data. As far as the kind of responses, we will devote this work to the study of repeated count and binary data.

1.1 Generalized linear fixed and mixed models

1.1.1 Generalized linear models (GLMs)

Let $\{y_i\}, i = 1, \cdots, K$, denote the observed independent responses, and $\mathbf{x}_i = (x_{i1}, \cdots, x_{ip})^\top$ be the associated p-dimensional covariate vector, whose effects on the response mean

$\mu_i(\boldsymbol{\beta}) = \mathrm{E}(Y_i)$ are given through a linear predictor $\mathbf{x}_i^\top \boldsymbol{\beta}$ with $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$. In the GLM regression setup (Nelder and Wedderburn, 1972), the responses y_i's are further assumed to follow the exponential family density function

$$f(y_i|\theta_i) = \exp[y_i\theta_i - a(\theta_i) + b(y_i)], \tag{1.1}$$

with the functional form of $a(\cdot)$ known, and $b(\cdot)$ depending only on y_i. Then it can be shown that the mean and variance functions of the response variable Y_i for all $i = 1, \cdots, K$, are given by

$$\begin{aligned} \mu_i(\boldsymbol{\beta}) &= \mathrm{E}(Y_i|\mathbf{x}_i) = a'(\theta_i), \text{ and} \\ \sigma_{ii}(\boldsymbol{\beta}) &= \mathrm{var}(Y_i|\mathbf{x}_i) = a''(\theta_i), \end{aligned} \tag{1.2}$$

respectively, where $a'(\cdot)$ and $a''(\cdot)$ are respectively the first and second derivatives of $a(\cdot)$ with respect to θ_i. The mean μ_i is related to the linear predictor $\mathbf{x}_i^\top \boldsymbol{\beta}$ by the link function $h(\cdot)$ as

$$h(\mu_i(\boldsymbol{\beta})) = \mathbf{x}_i^\top \boldsymbol{\beta}, \tag{1.3}$$

and $\theta_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ when the link function is canonical (McCullagh and Nelder, 1989).

1.1.1.1 Quasi-likelihood estimation for $\boldsymbol{\beta}$

In the above exponential family setup, $\hat{\boldsymbol{\beta}}$, the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ whenever $\boldsymbol{\beta}$ lies in an open subset in real space, is obtained by solving the equation:

$$\sum_i [y_i - a'(\theta_i)] \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} = \sum_i \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} \frac{a''(\theta_i)}{a''(\theta_i)} [y_i - a'(\theta_i)]$$

$$= \sum_i \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} \frac{\partial \mu_i}{\partial \theta_i} \left[\sigma_{ii}(\boldsymbol{\beta})\right]^{-1} \left[y_i - a'(\theta_i)\right] = 0, \tag{1.4}$$

which further gives the quasi-likelihood (QL) estimating equation

$$\sum_i \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \left[\sigma_{ii}(\boldsymbol{\beta})\right]^{-1} \left[y_i - \mu_i(\boldsymbol{\beta})\right] = 0 \tag{1.5}$$

proposed by Wedderburn (1974) (see also McCullagh, 1983, McCullagh and Nelder, 1989). Note that when applying QL estimating equation (1.5), one needs to specify only the first two moments of the distribution of Y_i's, even when the exact form of the distribution of Y_i's is unknown. It is known that this QL estimator $\hat{\boldsymbol{\beta}}_{QL}$ is a consistent estimation of true $\boldsymbol{\beta}$. For Poisson and binary data, whose distributions belong to the exponential family, $\hat{\boldsymbol{\beta}}_{QL}$ from (1.5) is also the ML estimate from (1.4).

1.1.2 Generalized linear mixed models (GLMMs)

In GLM, from (1.3), the mean μ_i is a function of the linear predictor $\mathbf{x}_i^\top \boldsymbol{\beta}$, which can be denoted as $\mu_i(\boldsymbol{\beta}) = h^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta}) = g(\mathbf{x}_i^\top \boldsymbol{\beta})$. If the responses are also affected by a latent random effect, the random effect can be included in the model through the linear predictor as $\mathbf{x}_i^\top \boldsymbol{\beta} + \tau_i^*$, where $\tau_i^* = \sigma_\tau \tau_i$ is an i.i.d. (independent and identically distributed) random variable with mean 0 and variance σ_τ^2. Then the GLMM can be defined by the conditional mean of Y_i given τ_i as

$$\mathrm{E}(Y_i|\mathbf{x}_i, \tau_i) = g(\mathbf{x}_i^\top \boldsymbol{\beta} + \sigma_\tau \tau_i), i = 1, \cdots, K. \tag{1.6}$$

In practice, τ_i^*'s are usually assumed to follow a normal or t distribution (Breslow and Clayton, 1993, Jiang, 1998), but there also exist some studies avoiding distributional assumption for τ_i^* (Montalvo, 1997, Wooldridge, 1999). As compared to the GLMs

(1.3), it is of main interest here for (1.6) to estimate σ_τ^2 in additional to $\boldsymbol{\beta}$. For inferences about $\boldsymbol{\beta}$ and σ_τ^2 under GLMs for longitudinal data, for example, we refer to Breslow and Clayton (1993), Jiang (1998), and Sutradhar (2004).

1.2 Semi-parametric GLMs (SGLMs)

In semi-parametric problems, the response y_i is influenced by both primary covariate $\mathbf{x}_i$ and certain secondary covariate z_i. As a result, in semi-parametric models, the mean response μ_i should be a function of both the fixed regression effect parameter $\boldsymbol{\beta}$, and an unspecified (nonparametric) function $\psi(z_i)$ that we assume to be smooth enough. That is, under semi-parametric setup, the mean response can be abbreviated as

$$\mu_i(\boldsymbol{\beta}, \psi(z_i)) = E(Y_i|\mathbf{x}_i, z_i) = g(\mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z_i)). \tag{1.7}$$

It is clear that when z_i is assumed to influence y_i through (1.7), any estimate obtained for $\boldsymbol{\beta}$ by ignoring $\psi(z_i)$ would be biased and hence mean squared error inconsistent. As compared to the parametric GLMs, the semi-parametric GLMs allow a more flexible treatment of the effects from the secondary covariate z_i.

In a semi-parametric setup, the fixed regression parameter vector $\boldsymbol{\beta}$ as well as the nonparametric function $\psi(\cdot)$ need to be estimated, even though our primary interest is only on $\boldsymbol{\beta}$. The $\boldsymbol{\beta}$ estimation approach presented in Section 1.1 was developed through the research on the parametric GLMs. Similarly, there exist many early works (Muller, 1988, Staniswalis, 1989) on nonparametric models somewhat equivalent to substituting $\boldsymbol{\beta} = 0$ in (1.7), yielding many kernel methods and their variants for nonparametric regression estimation, such as Nadaraya-Watson kernel regression estimation (Nadaraya, 1964, Watson, 1964, Bierens, 1987, Andrews, 1995), local linear and polynomial regression (Cleveland, 1979, Fan, 1992, 1993, Stone, 1980, 1982), recursive kernel estimation (see e.g., Ahmad and Lin, 1976, Greblicki and Krzyzak, 1980), spline smoothing (Whittaker, 1922, Eubank, 1988, Wahba, 1990), and nearest neighbor estimation (Royall, 1966, Stone, 1977). Among these techniques, the Nadaraya-Watson kernel estimator or the local constant estimator for $\psi(z)$ is the simplest to implement, and serves our purpose in this work well. As an illustrative example, in the nonparametric regression model

$$y_i = \psi(z_i) + \epsilon_i, i = 1, \cdots, K, \text{ and } \epsilon_i \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2),$$

this estimator at a given covariate point z is given by

$$\hat{\psi}(z) = \frac{\sum_{i=1}^{K} y_i K\left(\frac{z-z_i}{b}\right)}{\sum_{i=1}^{K} K\left(\frac{z-z_i}{b}\right)},$$

where $K(\cdot)$ is a suitable kernel density function and b is the bandwidth.

The performance of the kernel techniques is heavily influenced by the selection of an appropriate bandwidth parameter b, which is always a problem in nonparametric regression (Silverman, 1986). Many data-based procedures to choose a bandwidth, such as cross validation (see Stone, 1974, Picard and Cook, 1984, Kohn et al., 1991), generalized cross validation (Craven and Wahba, 1979) have been discussed in the literature. Altman (1990) suggested that these bandwidth selection techniques do not perform well when the errors are correlated, so we exclude these techniques from further discussion. Pagan and Ullah (1999) proposed an optimum value for bandwidth, which minimizes the approximate mean integrated squared error. The authors recommended $b \propto K^{-1/5}$, and suggested that this value of bandwidth is the only choice for b where the bias and variance, when estimating the model parameters, are of the same order of magnitude. In practice, the bias and variance cannot be minimized together by certain b value, so the best choice of b involves a trade-off between bias and variance (Ruppert, 1997). In this regard, based on the asymptotic formula for the nonparametric function estimators, we developed a mini-max approach, which selects the b values minimizing the maximum mean squared error of the estimate over the support of nonparametric functions.

In semi-parametric setup for independent responses, the estimation of both $\boldsymbol{\beta}$ and $\psi(\cdot)$ is also extensively studied in the literature (e.g., Severini and Staniswalis, 1994, Carota and Parmigiani, 2002). Based on the QL method, for example, Severini and Staniswalis (1994) proposed a semi-parametric QL (SQL) approach for the estimation of $\boldsymbol{\beta}$ and $\psi(\cdot)$. In this approach, one only needs to specify the form of the conditional mean $\mu_i = \mathrm{E}(Y_i|\mathbf{x}_i, z_i)$ as a function of $\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i)$, and the conditional variance $\sigma_{ii} = \mathrm{var}(Y_i|\mathbf{x}_i, z_i)$ as a function of μ_i, without the need of knowing the distribution of data. Under the assumption that the individuals are independent, the SQL estimating equations for $\psi(z)$ and $\boldsymbol{\beta}$ estimation can be written out as

$$\sum_{i=1}^{K} w_i(z) \frac{\partial \mu_i(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))}{\partial \psi(z)} \sigma_{ii}^{-1}(\mu_i(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))) \left[y_i - \mu_i(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))\right] = 0 \qquad (1.8)$$

(for all z on the support of $\psi(\cdot)$), and

$$\sum_{i=1}^{K} \frac{\partial \mu_i(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i))}{\partial \boldsymbol{\beta}} \sigma_{ii}^{-1}(\mu_i(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i))) \left[y_i - \mu_i(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i))\right] = 0, \qquad (1.9)$$

respectively. Here $w_i(z) = \frac{p_i\left(\frac{z-z_i}{b}\right)}{\sum_{i=1}^{K} p_i\left(\frac{z-z_i}{b}\right)}$, with $p_i(\cdot)$ being a kernel density function for which, for example, one may choose $p_i\left(\frac{z-z_i}{b}\right) = \frac{1}{\sqrt{2\pi}b}\exp\left(-\frac{1}{2}\left(\frac{z-z_i}{b}\right)^2\right)$ with a suitable bandwidth b. Note that when $w_i(z) = 1$, this SQL equation further reduces to the well-known quasi-likelihood estimating equation (Wedderburn, 1974). The authors applied their estimation methodology to continuous linear and gamma data, and discrete binary data. Note that in this thesis, we focus on only semi-parametric modeling and inferences of longitudinal discrete such as count and binary data, where independent count and binary data are special cases. As a preparation for inducing the estimation approaches in more general longitudinal setup, we now explain semi-parametric QL estimation in details for linear, count data and binary data models in the independence setup.

1.2.1 Linear model

The semi-parametric linear model can be written as

$$y_i = \mu_i(\boldsymbol{\beta}, \psi(z_i)) + \epsilon_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z_i) + \epsilon_i, i = 1, \cdots, K, \text{ and } \epsilon_i \overset{i.i.d.}{\sim} (0, \sigma_\epsilon^2). \quad (1.10)$$

Here $\mathrm{E}(Y_i|\mathbf{x}_i, z_i) = \mu_i(\boldsymbol{\beta}, \psi(z_i)) = \mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z_i)$ and $\mathrm{var}(Y_i|\mathbf{x}_i, z_i) = \sigma_{ii} = \sigma_\epsilon^2, i = 1, \cdots, K$. If ϵ_i's are normally distributed, the canonical link function $h(\cdot)$ is the identity function, and the natural parameter $\theta_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z_i)$. However, for applying the SQL approach, only mean and variance need to be specified, while the exact form of the distribution is irrelevant.

1.2.1.1 Estimation of nonparametric function $\psi(z)$

Whenever $\mu_i(\boldsymbol{\beta}, \psi(z_i))$ and $\sigma_{ii}(\mu_i)$ are correctly defined, the SQL estimating equations can be obtained directly from (1.8) and (1.9). For this model, according to (1.8), the SQL estimating equation for $\psi(z)$ is

$$\sum_{i=1}^{K} w_i(z) \frac{\partial \mu_i(\boldsymbol{\beta}, \psi(z))}{\partial \psi(z)} \left[\frac{y_i - \mu_i(\boldsymbol{\beta}, \psi(z))}{\sigma_\epsilon^2} \right] = 0 \quad (1.11)$$

for all z in the support of $\psi(\cdot)$. Because $\frac{\partial \mu_i(\boldsymbol{\beta}, \psi(z))}{\partial \psi(z)} = \frac{\partial [\mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z)]}{\partial \psi(z)} = 1$, (1.11) can be simplified as

$$\sum_{i=1}^{K} w_i(z) \left[\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \psi(z)}{\sigma_\epsilon^2} \right] = 0 \tag{1.12}$$
$$\Rightarrow \sum_{i=1}^{K} w_i(z) \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right) - \sum_{i=1}^{K} w_i(z) \psi(z) = 0,$$

yielding an estimate for the nonparametric function $\psi(z)$ as

$$\hat{\psi}(z) = \frac{\sum_{i=1}^{K} w_i(z) \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right)}{\sum_{i=1}^{K} w_i(z)} = \sum_{i=1}^{K} w_i(z) \left(y_i - \mathbf{x}_i^\top \boldsymbol{\beta} \right) \tag{1.13}$$

since $\sum_{i=1}^{K} w_i(z) = 1$. (1.13) is a general formula for any point z in the support of the nonparametric function $\psi(\cdot)$. Specifically at point z_i, the ith observed "secondary" covariate, it becomes

$$\hat{\psi}(z_i) = \sum_{j=1}^{K} w_j(z_i) \left(y_j - \mathbf{x}_j^\top \boldsymbol{\beta} \right) = \hat{y}_i - \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}, \tag{1.14}$$

where

$$\hat{y}_i = \sum_{j=1}^{K} w_j(z_i) y_j \qquad \text{and} \qquad \hat{\mathbf{x}}_i = \sum_{j=1}^{K} w_j(z_i) \mathbf{x}_j \,. \tag{1.15}$$

Note that the regression parameter vector $\boldsymbol{\beta}$ in (1.14) is unknown either, and need to be estimated from (1.9). In practice, the estimates of $\psi(z)$ and $\boldsymbol{\beta}$ are obtained by solving (1.8) and (1.9) iteratively until they both converge. The estimating equation for $\boldsymbol{\beta}$ under the present semi-parametric linear model is provided in the following section, although, these formulas for $\hat{\psi}(z_i)$ and $\hat{\boldsymbol{\beta}}$ are already discussed in literature. See Severini and Staniswalis (1994), Speckman (1988), Hastie and Tibshirani (1990).

1.2.1.2 Estimation of regression effects $\boldsymbol{\beta}$

In semi-parametric problem, the estimator $\hat{\psi}(\cdot)$ of the nonparametric function $\psi(\cdot)$ is also a function of unknown parameter vector $\boldsymbol{\beta}$. So after we substitute $\hat{\psi}(\cdot)$ for $\psi(\cdot)$ in (1.9) to write out SQL estimating equation for $\boldsymbol{\beta}$, the derivative of μ_i with respect to $\boldsymbol{\beta}$ need to take $\boldsymbol{\beta}$ in $\hat{\psi}(\cdot)$ into account. As for the present semi-parametric linear model, we first write $\mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i)) = \mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i)$ and compute

$$\begin{aligned}\frac{\partial \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i))}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i)\right] = \frac{\partial}{\partial \boldsymbol{\beta}} \left[\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{y}_i - \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}\right] \\ &= \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top, \end{aligned} \tag{1.16}$$

where $\hat{\mathbf{x}}_i$ is defined in (1.15). Then from (1.9) we can write the SQL estimating equation for $\boldsymbol{\beta}$ as

$$\sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left[\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \hat{\psi}(z_i)}{\sigma_\epsilon^2}\right] = 0,$$

and by substituting $\hat{\psi}(z_i) = \hat{y}_i - \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}$ we obtain

$$\sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left[y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \hat{y}_i + \hat{\mathbf{x}}_i^\top \boldsymbol{\beta}\right] = \sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left[\left(y_i - \hat{y}_i\right) - \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \boldsymbol{\beta}\right] = 0,$$

yielding

$$\sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left(y_i - \hat{y}_i\right) = \sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right) \boldsymbol{\beta}.$$

It then follows that $\hat{\boldsymbol{\beta}}$ has the closed form expression given by

$$\hat{\boldsymbol{\beta}} = \left[\sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)\right]^{-1} \sum_{i=1}^{K} \left(\mathbf{x}_i - \hat{\mathbf{x}}_i\right)^\top \left(y_i - \hat{y}_i\right), \tag{1.17}$$

where $\hat{y}_i$ and $\hat{\mathbf{x}}_i$ are given in equation (1.15). The above equation (1.17) is the same as in Severini and Staniswalis (1994) [Eq. (10), page 503] with $D = I$, the identity matrix.

1.2.2 Count data model

The ideal case for count data is to follow Poisson density function $f(y_i)$, which can be expressed as a special form of exponential family density (1.1) given by

$$f(y_i) = \frac{\exp(-\mu_i)\mu_i^{y_i}}{y_i!} = \frac{1}{y_i!}\exp[y_i \log \mu_i - \mu_i], \tag{1.18}$$

where $\theta_i = \log \mu_i$ and $a(\theta_i) = \mu_i$.

Thus we write the Poisson mean and variance as

$$\mathrm{E}(Y_i|\mathbf{x}_i, z_i) = \mathrm{var}(Y_i|\mathbf{x}_i, z_i) = \mu_i(\boldsymbol{\beta}, \psi(z_i)), \tag{1.19}$$

where

$$\mu_i(\boldsymbol{\beta}, \psi(z_i)) = \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z_i)),$$

which is different from (1.10) under the linear case, but in the present semi-parametric setup still consists of the fixed regression function as well as a nonparametric smooth function.

In practice, count data seldom exactly follow Poisson distribution, but usually their first two moments are still modeled by (1.19). Because SQL requires only the correct specification of mean and variance, SQL approach can thus be applied to real count data.

1.2.2.1 Estimation of nonparametric function $\psi(z)$

For constructing SQL estimating equation for $\psi(z)$, we first compute $\frac{\partial \mu_i(\boldsymbol{\beta}, \psi(z))}{\partial \psi(z)} = \frac{\partial \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z))}{\partial \psi(z)} = \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z))$. Then by substituting this result as well as the formulas (1.19) for mean and variance into (1.8), we obtain the SQL estimating equation for $\psi(z)$ as

$$\sum_{i=1}^{K} w_i(z) \left[y_i - \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \psi(z)) \right] = 0 \tag{1.20}$$

for all z in the support of $\psi(\cdot)$, which further gives a closed form solution

$$\hat{\psi}(z) = \log \left(\frac{\sum_{i=1}^{K} w_i(z) y_i}{\sum_{i=1}^{K} w_i(z) \exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \right).$$

Thus for $z = z_i$ the estimator of $\psi(z)$ has the form

$$\hat{\psi}(z_i) = \log \left(\frac{\sum_{j=1}^{K} w_j(z_i) y_j}{\sum_{j=1}^{K} w_j(z_i) \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right). \tag{1.21}$$

1.2.2.2 Estimation of regression effects $\boldsymbol{\beta}$

For establishing the SQL estimating equation for $\boldsymbol{\beta}$, we first need to compute

$$\frac{\partial \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i))}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i)) = \left[\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i)) \right] \left[\mathbf{x}_i + \frac{\partial \hat{\psi}(z_i)}{\partial \boldsymbol{\beta}} \right] \tag{1.22}$$

with $\hat{\psi}(z_i)$ as in (1.21). The derivative $\frac{\partial \hat{\psi}(z_i)}{\partial \boldsymbol{\beta}}$ is computed as

$$\begin{aligned} \frac{\partial \hat{\psi}(z_i)}{\partial \boldsymbol{\beta}} &= - \left[\frac{\sum_{j=1}^{K} w_j(z_i) y_j}{\sum_{j=1}^{K} w_j(z_i) \exp(\mathbf{x}_j^\top \boldsymbol{\beta})} \right]^{-1} \frac{\left[\sum_{j=1}^{K} w_j(z_i) y_j \right] \left[\sum_{j=1}^{K} w_j(z_i) \exp(\mathbf{x}_j^\top \boldsymbol{\beta}) \mathbf{x}_j \right]}{\left[\sum_{j=1}^{K} w_j(z_i) \exp(\mathbf{x}_j^\top \boldsymbol{\beta}) \right]^2} \\ &= - \frac{\sum_{j=1}^{K} w_j(z_i) \exp(\mathbf{x}_j^\top \boldsymbol{\beta}) \mathbf{x}_j}{\sum_{j=1}^{K} w_j(z_i) \exp(\mathbf{x}_j^\top \boldsymbol{\beta})}. \end{aligned} \tag{1.23}$$

Now by using (1.23) in (1.22) we write

$$\begin{aligned}\frac{\partial \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i))}{\partial \boldsymbol{\beta}} &= \left[\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i))\right] \left[\mathbf{x}_i - \frac{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})\mathbf{x}_j}{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}\right] \\ &= \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i)) \left[\mathbf{x}_i - \frac{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})\mathbf{x}_j}{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}\right].\end{aligned}$$

Then by substituting all these results into (1.9), the SQL estimating equation for $\boldsymbol{\beta}$ for count data is obtained as

$$\sum_{i=1}^K \left[\mathbf{x}_i - \frac{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})\mathbf{x}_j}{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}\right] [y_i - \tilde{\mu}_i] = 0,$$

where $\tilde{\mu}_i = \exp(\mathbf{x}_j^\top \boldsymbol{\beta} + \hat{\psi}(z_i))$. Now by using

$$\hat{\mathbf{x}}_i = \frac{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})\mathbf{x}_j}{\sum_{j=1}^K w_j(z_i)\exp(\mathbf{x}_j^\top \boldsymbol{\beta})}, \tag{1.24}$$

we rewrite the estimating equation as

$$\sum_{i=1}^K (\mathbf{x}_i - \hat{\mathbf{x}}_i)^\top (y_i - \tilde{\mu}_i) = 0. \tag{1.25}$$

The estimating equation (1.25) can be solved iteratively using the well-known Newton-Raphson method. The iterative equation has the form

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{(r+1)} &= \hat{\boldsymbol{\beta}}_{(r)} - \left[\frac{\partial}{\partial \boldsymbol{\beta}^\top} \sum_{i=1}^K (\mathbf{x}_i - \hat{\mathbf{x}}_i)^\top (y_i - \tilde{\mu}_i)\right]^{-1} \left[\sum_{i=1}^K (\mathbf{x}_i - \hat{\mathbf{x}}_i)(y_i - \tilde{\mu}_i)\right] \\ &= \hat{\boldsymbol{\beta}}_{(r)} + \left[\sum_{i=1}^K (\mathbf{x}_i - \hat{\mathbf{x}}_i)^\top \tilde{\mu}_i (\mathbf{x}_i - \hat{\mathbf{x}}_i)\right]^{-1} \left[\sum_{i=1}^K (\mathbf{x}_i - \hat{\mathbf{x}}_i)(y_i - \tilde{\mu}_i)\right]\end{aligned} \tag{1.26}$$

and is used to compute the final estimate $\hat{\boldsymbol{\beta}}$ until convergence.

1.2.3 Binary data model

Unlike linear and count data whose distribution and variances are usually not known, the binary distribution $f(y_i)$ and variance $\sigma_{ii} = \text{var}(Y_i|\mathbf{x}_i, z_i)$ can always be written out with mean μ_i as

$$\begin{aligned} f(y_i) &= \mu_i^{y_i}(1-\mu_i)^{1-y_i} \text{ and} \\ \sigma_{ii} &= \mu_i(\boldsymbol{\beta}, \psi(z_i))\left[1-\mu_i(\boldsymbol{\beta}, \psi(z_i))\right] \end{aligned} \tag{1.27}$$

respectively, in the semi-parametric GLM setup for binary responses. The binary density is a special case of the exponential family density (1.1) with

$$\theta_i = \log\left(\frac{\mu_i}{1-\mu_i}\right) \text{ and } a(\theta_i) = -\log(1-\mu_i).$$

Now for implementing the SQL estimation approach, we only need to specify the model for conditional mean μ_i. Under the canonical link function, the canonical parameter $\theta_i = \mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i)$, $a(\theta_i) = \log(1+\exp(\theta_i))$, and

$$\mu_i = \frac{\exp(\theta_i)}{1+\exp(\theta_i)},$$

yielding

$$\text{E}(Y_i|\mathbf{x}_i, z_i) = a'(\theta_i) = \mu_i(\boldsymbol{\beta}, \psi(z_i)) = \frac{\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i))}{1+\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z_i))}.$$

1.2.3.1 Estimation of nonparametric function $\psi(z)$

As usual, we first compute

$$\frac{\partial \mu_i(\boldsymbol{\beta}, \psi(z))}{\partial \psi(z)} = \frac{\partial}{\partial \psi(z)} \frac{\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))}{1+\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))} = \frac{\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))}{1+\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))} \frac{1}{1+\exp(\mathbf{x}_i^\top\boldsymbol{\beta} + \psi(z))}$$

$$= \mu_i(\boldsymbol{\beta}, \psi(z)) \left[1 - \mu_i(\boldsymbol{\beta}, \psi(z))\right],$$

and then simplify (1.8) with these results as

$$\sum_{i=1}^{K} w_i(z) \left[y_i - \mu_i(\boldsymbol{\beta}, \psi(z))\right] = 0 \tag{1.28}$$

for all z in the support of $\psi(\cdot)$, which is the SQL estimating equation for $\psi(z)$. Note that it has the same form as (1.20) with the difference lying in the formula for $\mu_i(\boldsymbol{\beta}, \psi(z))$.

1.2.3.2 Estimation of regression effects $\boldsymbol{\beta}$

In binary case, there is no closed form solution for $\hat{\psi}(z)$. For computing $\frac{\partial \hat{\psi}(z)}{\partial \boldsymbol{\beta}}$, we replace $\psi(\cdot)$ with $\hat{\psi}(\cdot)$ in (1.28), and then take derivative of both sides with respect to $\boldsymbol{\beta}$ to obtain

$$-\sum_{i=1}^{K} w_i(z)\mu_i(\boldsymbol{\beta}, \hat{\psi}(z)) \left[1 - \mu_i(\boldsymbol{\beta}, \hat{\psi}(z))\right] \left[\mathbf{x}_i + \frac{\partial \hat{\psi}(z)}{\partial \boldsymbol{\beta}}\right] = 0\,.$$

By solving for $\frac{\partial \hat{\psi}(z)}{\partial \boldsymbol{\beta}}$, we obtain

$$\frac{\partial \hat{\psi}(z)}{\partial \boldsymbol{\beta}} = -\frac{\sum_{i=1}^{K} w_i(z)\mu_i(\boldsymbol{\beta}, \hat{\psi}(z)) \left[1 - \mu_i(\boldsymbol{\beta}, \hat{\psi}(z))\right] \mathbf{x}_i}{\sum_{i=1}^{K} w_i(z)\mu_i(\boldsymbol{\beta}, \hat{\psi}(z)) \left[1 - \mu_i(\boldsymbol{\beta}, \hat{\psi}(z))\right]}.$$

Then

$$\begin{aligned}
\frac{\partial \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i))}{\partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left[\frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i))}{1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i))}\right] \\
&= \left[\frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i))}{\left[1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta} + \hat{\psi}(z_i))\right]^2}\right] \left[\mathbf{x}_i + \frac{\partial \hat{\psi}(z_i)}{\partial \boldsymbol{\beta}}\right]
\end{aligned}$$

$$= \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i)) \left[1 - \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i))\right] \left[\mathbf{x}_i + \frac{\partial \hat{\psi}(z_i)}{\partial \boldsymbol{\beta}}\right].$$

Based on all these results, the SQL estimating equation (1.9) reduces to

$$\sum_{i=1}^{K} \left[\mathbf{x}_i + \frac{\partial \hat{\psi}(z_i)}{\partial \boldsymbol{\beta}}\right] \left[y_i - \mu_i(\boldsymbol{\beta}, \hat{\psi}(z_i))\right] = 0, \tag{1.29}$$

which is the SQL estimating equation for $\boldsymbol{\beta}$ in binary case, and need to be solved iteratively using the Newton-Raphson method.

1.3 Generalized linear fixed and mixed models for longitudinal data

1.3.1 Generalized linear fixed models for longitudinal data

The results on GLMs in Section 1.1 and semi-parametric GLMs in Section 1.2 were based on independent observations. In this section, we provide an overview of the existing models and associated inferences in a longitudinal setup.

In notation, for the ith ($i = 1, \cdots, K$) individual, let $\mathbf{y}_i = (y_{i1}, \cdots, y_{ij}, \cdots, y_{in_i})^\top$ denote $n_i \times 1$ vector of repeated responses, where y_{ij} is the response recorded at time j. Further, suppose that y_{ij} is influenced by a fixed and known p-dimensional time dependent covariate vector $\mathbf{x}_{ij} = (x_{ij1}, \cdots, x_{ijv}, \cdots, x_{ijp})^\top$ collected together with y_{ij}, and the regression effects of $\mathbf{x}_{ij}$ on y_{ij} for all $i = 1, \cdots, K$ and $j = 1, \cdots, n_i$ can be indicated by a p-dimensional vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$. Because the same variables are measured repeatedly on the same individual over a period of time, the observations are likely to be correlated. So the correlations among y_{ij}'s for the same ith individual cannot be neglected, even though it may still be reasonable to assume independence among different individuals. The joint distribution of response vector $\mathbf{y}_i$ is hard to determine, especially in discrete cases, but we assume that, conditional on the covariates, each component y_{ij} marginally follows (1.1), and has mean $\mu_{ij}(\boldsymbol{\beta}) = \mathrm{E}[Y_{ij}] = a'(\theta_{ij})$ and variance $\sigma_{ijj}(\boldsymbol{\beta}) = \mathrm{var}[Y_{ij}] = a''(\theta_{ij})$ [see (1.2)–(1.3)]. Following (1.5), the independence assumption based QL estimating equation for the unknown regression parameter $\boldsymbol{\beta}$ can be written as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial a'(\theta_{ij})}{\partial\boldsymbol{\beta}}\left[a''(\theta_{ij})\right]^{-1}\left[y_{ij}-a'(\theta_{ij})\right] =$$
$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial \mu_{ij}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\left[\sigma_{ijj}(\boldsymbol{\beta})\right]^{-1}\left[y_{ij}-\mu_{ij}(\boldsymbol{\beta})\right] = 0\,. \tag{1.30}$$

$\boldsymbol{\beta}$ estimates from (1.30) are consistent. However, because the correlations among the observations from the same individual are ignored, such estimates are in general inefficient. In order to achieve the desired efficiency, it is necessary to take into account the correlations of longitudinal responses.

One of the first remedies to the inefficient estimation problem in longitudinal data analysis was proposed by Liang and Zeger (1986). These authors introduced a 'working' correlation matrix to account for the correlation among the repeated observations in the longitudinal setup, and proposed a generalized estimating equation (GEE) of the form

$$\sum_{i=1}^{K}\frac{\partial \boldsymbol{\mu}_i^\top}{\partial\boldsymbol{\beta}}\mathbf{V}_i^{-1}(\alpha)\left[\mathbf{y}_i-\boldsymbol{\mu}_i\right] = 0 \tag{1.31}$$

to obtain consistent and efficient regression estimates of the parameters involved in the GLM model for longitudinal data. Define $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}),\cdots,\mu_{ij}(\boldsymbol{\beta}),\cdots,\mu_{in}(\boldsymbol{\beta}))^\top$ as the mean vector of $\mathbf{y}_i$, and $\mathbf{V}_i(\alpha) = \mathbf{A}_i^{1/2}\mathbf{R}_i(\alpha)\mathbf{A}_i^{1/2}$ as the covariance matrix of $\mathbf{y}_i$ with $\mathbf{A}_i = \mathrm{diag}[\sigma_{i11}(\boldsymbol{\beta}),\cdots,\sigma_{ijj}(\boldsymbol{\beta}),\cdots,\sigma_{inn}(\boldsymbol{\beta})]$. Then, $\mathbf{R}_i(\alpha)$ is the 'working' correlation matrix with α as its 'working' correlation parameter. This GEE represented an important progress in longitudinal data analysis. However, subsequent research showed that it can fail to ensure consistency and efficiency in some situations. For example, Crowder (1995) showed that due to a problem in estimating the so-called 'working' correlation parameter α, the GEE regression parameter estimates are inconsistent in several situations. In cases where 'working' correlations are estimable, Sutradhar and Das (1999) demonstrated that the use of stationary 'working' correlation matrix in GEE can produce less efficient regression estimates than the independence assumption based QL or moment estimates. Sutradhar (2010) further demonstrated that even in the stationary setup, the use of 'working' stationary correlation matrix can still produce less efficient estimates than the 'working' independence assumption based GEE or QL, or moments estimates. Sutradhar (2003) proposed a generalization of the QL estimation approach, where $\boldsymbol{\beta}$ is obtained by solving the generalized quasi-likelihood (GQL) estimating equation given by

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i^\top}{\partial \boldsymbol{\beta}} \boldsymbol{\Sigma}_i^{-1}(\rho)\left[\mathbf{y}_i - \boldsymbol{\mu}_i\right] = 0\,, \tag{1.32}$$

where $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\mu_{i1}(\boldsymbol{\beta}), \cdots, \mu_{ij}(\boldsymbol{\beta}), \cdots, \mu_{in_i}(\boldsymbol{\beta}))^\top$ is the mean vector of $\mathbf{y}_i$, $\boldsymbol{\Sigma}_i(\rho) = \mathbf{A}_i^{1/2}\mathbf{C}_i(\rho)\mathbf{A}_i^{1/2}$ is the covariance matrix of $\mathbf{y}_i$ with $\mathbf{A}_i = \mathrm{diag}[\sigma_{i11}(\boldsymbol{\beta}), \cdots, \sigma_{ijj}(\boldsymbol{\beta}), \cdots, \sigma_{in_in_i}(\boldsymbol{\beta})]$, $\mathbf{C}_i(\rho)$ is a general class of auto-correlations, and ρ is a correlation index parameter. Note that GQL allows each individual to have different number of repeated responses, n_i. The estimator $\hat{\boldsymbol{\beta}}_{GQL}$ obtained by solving (1.32) is consistent and very efficient for $\boldsymbol{\beta}$.

1.3.2 Generalized linear mixed models for longitudinal data

For a generalization from GLM to GLMM for longitudinal data, we can follow the procedure presented in Section 1.1.2, that is, to add a random effect to the linear predictor in the mean function, and thus convert the mean in GLM to the conditional mean in GLMM for longitudinal data. To be specific, suppose that the mean of Y_{ij} in GLM is given by $\mu_{ij}(\boldsymbol{\beta}) = \mathrm{E}(Y_{ij}) = g(\mathbf{x}_{ij}^\top \boldsymbol{\beta})$, then in GLMM, the conditional mean of Y_{ij} given τ_i, the random effect as defined in (1.6), is given by

$$\mathrm{E}(Y_{ij}|\mathbf{x}_{ij}, \tau_i) = g(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \sigma_\tau \tau_i). \tag{1.33}$$

In this longitudinal setup, the repeated responses of the same individual share a common random effect, which will influence the correlation structure of the model. Extra efforts are required for defining the dynamic dependence of the repeated responses conditional on random effect τ_i, and computing the unconditional correlation structure of the model. For count data, for example, such conditional dynamic dependence can be

$$y_{ij}|\tau_i \;=\; \sum_{k=1}^{y_{i,j-1}} b_k(\rho)|\tau_i + d_{ij}|\tau_i, \quad j = 2, \ldots, n_i \tag{1.34}$$

(Sutradhar and Bari, 2007), where it is assumed that $y_{i1}|\tau_i \sim \mathrm{Poi}(m_{i1}^*)$, and for $j = 2, \ldots, n_i$, $y_{i,j-1}|\tau_i \sim \mathrm{Poi}(m_{i,j-1}^*)$, and $d_{ij}|\tau_i \sim \mathrm{Poi}(m_{ij}^* - \rho m_{i,j-1}^*)$ with $m_{ij}^* = \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \sigma_\tau \tau_i)$ for $j = 1, \ldots, n_i$. Here $\mathrm{Poi}(m)$ stands for Poisson distribution with mean m. In (1.34), conditional on τ_i, d_{ij} and $y_{i,j-1}$ are independent. Furthermore, $b_k(\rho)$ stands for a binary random variable with $\Pr[b_k(\rho) = 1] = \rho$. Model (1.34) produces correlation structure reflecting longitudinal relationship for over-dispersed count data. Similarly, for longitudinal binary data, the GLMM, for example, has the form

$$\Pr(y_{ij}=1|y_{i,j-1},\mathbf{x}_{ij},\tau_i)=\frac{\exp(\mathbf{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta}+\theta y_{i,j-1}+\sigma_\tau\tau_i)}{1+\exp(\mathbf{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta}+\theta y_{i,j-1}+\sigma_\tau\tau_i)}, \quad \text{for } j=2,\cdots,n_i \tag{1.35}$$

(Sutradhar et al., 2008), where θ is a dynamic dependence parameter, and σ_τ is the random effect standard derivation.

For inferences under the count data model (1.34), we refer, for example, to Montalvo (1997), Wooldridge (1999), Jowaheer and Sutradhar (2002), Sutradhar and Bari (2007), Winkelmann (2008), and for inferences under the binary dynamic model (1.35), we refer, for example, to Manski (1987), Honoré and Kyriazidou (2000), Sutradhar et al. (2008, 2010).

1.4 Semi-parametric generalized linear fixed models for longitudinal data

The GLMs explained in Section 1.3.1 has also been generalized to a semi-parametric setup (Severini and Staniswalis, 1994, Lin and Carroll, 2001). Under this generalization, the mean and variance functions are defined as

$$\begin{aligned}\mu_{ij}(\boldsymbol{\beta},\psi(z_{ij})) &= \mathrm{E}(Y_{ij}|\mathbf{x}_{ij},\psi(z_{ij}))=a'(\theta_{ij}), \text{ and}\\ \sigma_{ijj}(\boldsymbol{\beta},\psi(z_{ij})) &= \mathrm{var}(Y_{ij}|\mathbf{x}_{ij},\psi(z_{ij}))=a''(\theta_{ij}),\end{aligned} \tag{1.36}$$

and the link function $h(\cdot)$ in (1.3) has the form

$$h(\mu_{ij}(\boldsymbol{\beta},\psi(z_{ij})))=\mathbf{x}_{ij}^{\top}\boldsymbol{\beta}+\psi(z_{ij}), \tag{1.37}$$

where $\psi(z_{ij})$ is the nonparametric function in secondary covariate z_{ij}.

For more studies for this type of SGLMs for longitudinal data, we refer to Zeger and Diggle (1994), Severini and Staniswalis (1994) (Section 8), Lin and Carroll (2001), Sneddon and Sutradhar (2004), You and Chen (2007), Warriyar and Sutradhar (2014), Sutradhar et al. (2016). In particular, in linear longitudinal setup, this type of model is studied by Severini and Wong (1992), Zeger and Diggle (1994), Moyeed and Diggle (1994), You and Chen (2007), Fan et al. (2007), Fan and Wu (2008), Li (2011), Warriyar and Sutradhar (2014).

1.4.1 Existing inferential techniques

Because this SGLM for longitudinal data is a generalization of the SGLM (1.7) to the longitudinal setup, it is convenient to use a general notation as follows. Suppose that t_{ij} denotes the time at which the jth ($j = 1, \ldots, n_i$) response is recorded from the ith ($i = 1, \ldots, K$) individual, and y_{ij} denotes this response. Next, unlike the scalar response case explained by model (1.7), suppose that $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ denotes the $n_i \times 1$ vector of repeated responses for the i-th ($i = 1, \ldots, K$) individual. Also suppose that y_{ij} is influenced by a fixed and known p-dimensional time-dependent primary covariate vector $\boldsymbol{x}_{ij}(t_{ij})$ and an additional time-dependent scalar secondary covariate $z_{ij}(t_{ij})$. Note that similar to (1.7), the primary covariates are included in the regression model parametrically using a linear predictor, whereas the covariate(s) of secondary interest are included in the model nonparametrically. Because of the fact that the repeated responses $\{y_{ij}, j = 1, \ldots, n_i\}$ are likely to be correlated, Severini and Staniswalis (1994) (Eq. (17)), and Lin and Carroll (2001) (Eq. (10)), for example, estimated the regression effects $\boldsymbol{\beta}$ by solving the so-called 'working' correlations-based

SGEE (semi-parametric generalized estimating equation)

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i^\top(\boldsymbol{\beta}, \boldsymbol{X}_i, \hat{\boldsymbol{\psi}}(\boldsymbol{\beta}, \boldsymbol{z}_i))}{\partial \boldsymbol{\beta}} \boldsymbol{V}_i^{-1} \left[\boldsymbol{y}_i - \boldsymbol{\mu}_i(\beta, X_i, \hat{\boldsymbol{\psi}}(\boldsymbol{\beta}, \boldsymbol{z}_i))\right] = 0, \tag{1.38}$$

where $\boldsymbol{V}_i$ is referred to as a 'working' covariance matrix. More specifically, in (1.38), $\boldsymbol{X}_i^\top = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})$ denotes the $p \times n_i$ covariate matrix with $\boldsymbol{x}_{ij}$ as the p-dimensional covariate vector for the i-th individual at time point t_{ij}, $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{X}_i, \hat{\boldsymbol{\psi}}(\boldsymbol{\beta}, \boldsymbol{z}_i))$ is a mean vector as opposed to the scalar mean $\mu_i(\cdot)$ in (1.9), and $\hat{\boldsymbol{\psi}}(\boldsymbol{\beta}, \boldsymbol{z}_i) = (\hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_{i1}), \cdots, \hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_{ij}), \cdots, \hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_{in_i}))^\top$ is an $n_i \times 1$ consistent estimate of the nonparametric vector function for known $\boldsymbol{\beta}$. Here $\boldsymbol{z}_i$ represents the secondary covariate values $z_{i1}, \ldots, z_{in_i}$. As far as $\boldsymbol{V}_i$ matrix is concerned, it is an $n_i \times n_i$ 'working' covariance matrix representing the correlations of the repeated responses. It is computed by

$$\boldsymbol{V}_i = \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{R}_i \boldsymbol{A}_i^{\frac{1}{2}}, \tag{1.39}$$

where $\boldsymbol{A}_i = \text{diag}[\text{var}(y_{i1}), \ldots, \text{var}(y_{in_i})]$ with $\text{var}(y_{ij}) = \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))$ for the Poisson panel data, and $\text{var}(y_{ij}) = \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))[1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))]$ for binary data, for examples. The matrix $\boldsymbol{R}_i$ has been computed by an unstructured (UNS) common constant correlation matrix ($\boldsymbol{R} = \boldsymbol{R}_i$), where

$$\boldsymbol{R} = K^{-1} \sum_{i=1}^{K} \boldsymbol{r}_i \boldsymbol{r}_i^\top, \text{ where } \boldsymbol{r}_i = (r_{i1}, \ldots, r_{in_i})^\top, \tag{1.40}$$

with $r_{ij} = [y_{ij} - \mu_{ij}(\boldsymbol{\beta}, x_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))]/[\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))]^{\frac{1}{2}}$ for count data, and $r_{ij} = [y_{ij} - \mu_{ij}(\boldsymbol{\beta}, x_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))]/[\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))[1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))]]^{\frac{1}{2}}$ for binary data, for examples. Severini and Staniswalis (1994) (Eq. (18)) and Lin and Carroll (2001) (Eqs. (6)-(7)) also estimated the nonparametric function $\psi(z)$ using the working correlation matrix $\boldsymbol{R}_i$, whereas Zeger and Diggle (1994) (see also Sneddon and Sutradhar, 2004, You and Chen, 2007) used 'working' independence among the repeated data.

However, the approach of Severini and Staniswalis (1994) and Lin and Carroll (2001), for the estimation of both $\boldsymbol{\beta}$ and the nonparametric function $\psi(\cdot)$, has several drawbacks:

1) The common matrix $\boldsymbol{R}$ can not be computed unless $n_i = n$ for all $i = 1, \ldots, K$. One cannot use this $\boldsymbol{R}$ matrix for panel data, especially when an $n_i \times n_i$ matrix is needed for the i-th individual (see Sutradhar, 2010). Furthermore, because covariates ($\boldsymbol{x}_{ij}$) of an individual i are dependent on j, Sutradhar (2010) showed that the correlations of the repeated data following a sensible dynamic model also involve $\boldsymbol{x}_{ij}$. This, for $j < k$, for a known function q, produces

$$\mathrm{E}[r_{ij} r_{ik}] = q(\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \hat{\psi}(z_{ij}), \hat{\psi}(z_{ik})) \tag{1.41}$$

and hence the average $K^{-1} \sum_{i=1}^{K} r_{ij} r_{ij}$ obtained from all individuals may be biased for the true correlation element $\rho_{i,jk}$ for the i-th individual. This will produce an inefficient estimate of $\boldsymbol{\beta}$, especially when the covariates are dependent on the value of j.

2) Weights are used to select data points with associated secondary covariate value z_{ij} close to the targeted point z for estimating nonparametric function value $\psi(z)$. Under these circumstances, if the correlations among repeated responses are considered, data points with different distances from point z will inevitably mix up, causing failure for the weights to select the correct data points for nonparametric function estimation. This is why using $\boldsymbol{R}_i$ matrices in weighted GEE, even the correct ones, for nonparametric function estimation can be counterproductive. In fact, Lin and Carroll (2001) (Section 7) found that using the $\boldsymbol{R}_i$ matrix for the estimation of $\psi(\cdot)$

produces a less efficient estimate than using the independence assumption, that is, $\boldsymbol{R}_i = \boldsymbol{I}_{n_i}$. Besides, $\psi(\cdot)$ is of secondary interest and hence it is sufficient to estimate it consistently, whereas more effort is needed to obtain a consistent and efficient estimate for the main regression parameter $\boldsymbol{\beta}$.

In fact, it is demonstrated in details in Chapter 4 (Section 4.1.4.1) that the above SGEE(UNS) approach produces less efficient regression estimates than independence assumption based such as SGEE(I) and SQL approaches. Thus, the SGEE(UNS) or generally speaking the SGEE approach may not be appropriate for such kind of problems.

1.4.2 A proposed inference remedy for the SGLFMs for longitudinal data

In this thesis, we will revisit the aforementioned inference issue. Specifically, under this semi-parametric longitudinal setup, we will discuss a SGQL (semi-parametric generalized quasi-likelihood) approach for consistent and efficient estimation of $\boldsymbol{\beta}$ for count data model (Chapter 2) and for binary data models (Chapter 4).

1.5 Objective of the thesis

The main objective of this thesis is to study the semi-parametric fixed and mixed models for discrete data, namely count and binary data. The plan of the thesis is as follows.

(i) As indicated above, in Chapter 2 we first revisit the fitting of a semi-parametric generalized linear fixed model to repeated count data. Specifically, we explain what was done by Sutradhar et al. (2016) in fitting such models. Both models and inference techniques used by these authors will be indicated.

(ii) In Chapter 3, we provide (a) a generalization of the semi-parametric fixed models for longitudinal count data discussed in Chapter 2 to the mixed model setup; (b) we discuss the consistent estimation of all functions and parameters involved in the model; here, we propose a SGQL approach for the estimation of the main regression and overdispersion parameters; (c) we show that this SGQL estimator is efficient through an intensive simulation study based on finite samples; and (d) the asymptotic properties of the estimators are presented.

(iii) Different from count data models, there exist several dynamic fixed models to deal with repeated binary data. The linear dynamic conditional probability (LDCP) and binary dynamic logit (BDL) models are widely used. (a) In Chapter 3, we provide a generalization of LDCP model to the semi-parametric setup. We discuss consistent estimation techniques for all functions and parameters of the model. The asymptotic and finite sample performances of the estimators are examined. (b) We illustrate the proposed semi-parametric model and estimation techniques by reanalyzing an infectious disease data set. Next, (c) we provide a generalization of the BDL model to the semi-parametric setup which is referred to as the SBDL model. Consistent and efficient estimates for functions and parameters are discussed both analytically and empirically.

(iv) In Chapter 5, (a) we further generalize the SBDFL (semi-parametric binary dynamic fixed logit) model discussed in Chapter 4 to the mixed model setup. This generalized model is referred to as the SBDML (semi-parametric binary dynamic mixed logit) model. Then (b) we provide consistent estimation of all functions and parameters involved in the model. Here we propose SGQL and SML (semi-parametric maximum likelihood) approaches for the estimation of the main regression and overdispersion parameters. (c) Asymptotic properties of the estimators are then derived. (d)

We also conduct a simulation study to compare the performances of the two approaches, namely the SGQL and SML approaches, for parameter estimation.

The conclusion of the thesis is given in Chapter 6.

Chapter 2

Semi-parametric dynamic fixed models for longitudinal count data

2.1 Semi-parametric dynamic model for panel count data

As indicated in Section 1.4.1, suppose that t_{ij} denotes the time at which the jth ($j = 1, \ldots, n_i$) response is recorded from the ith ($i = 1, \ldots, K$) individual, and y_{ij} denotes this response. Also suppose that $\boldsymbol{x}_{ij}(t_{ij})$ is the primary covariate collected at time t_{ij}, and $z_{ij}(t_{ij})$ is a secondary covariate corresponding to the same time t_{ij}. These z_{ij}'s are in general assumed to be dense. In some situations, one may be interested to know the direct influence of t_{ij} on the response y_{ij}. In such cases, $z_{ij}(t_{ij}) = t_{ij}$ (Lin and Carroll, 2001) where z_{ij} still retains its dense character. Because the effect of $z_{ij}(t_{ij})$ on y_{ij} is not of direct interest, its influence would be taken care of non-parametrically, whereas the effects of the primary covariates (those are of direct interest) $\boldsymbol{x}_{ij}(t_{ij})$ are formulated through a specified parametric regression function.

Next, in this longitudinal setup, the repeated responses $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ are likely to be correlated. In a longitudinal parametric setup, there exist many studies (Sutradhar, 2003, 2010, 2011), where the correlations are modeled through certain dynamic relationship between the present and past responses of the individual. This dynamic model has also been generalized recently by Sutradhar et al. (2016) to the longitudinal semi-parametric setup. More specifically, following Sutradhar (2003), these authors (Sutradhar et al., 2016) have used the dynamic model given by

$$\begin{cases} y_{i1} & = \mathrm{Poi}(\mu_{i1}(\boldsymbol{\beta}, \boldsymbol{x}_{i1}, \psi(z_{i1}))) \\ y_{ij} & = \rho * y_{i,j-1} + d_{ij} = \sum_{s=1}^{y_{i,j-1}} b_s(\rho) + d_{ij}, \;\; j = 2, \ldots, n_i, \end{cases} \tag{2.1}$$

where $\Pr[b_s(\rho) = 1] = \rho$ and $\Pr[b_s(\rho) = 0] = 1 - \rho$, with ρ as the correlation index parameter; and

$$d_{ij} \sim \mathrm{Poi}\left[\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) - \rho\mu_{i,j-1}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j-1}, \psi(z_{i,j-1}))\right]$$

for $j = 2, \ldots, n_i$, where in general for all $j = 1, \ldots, n_i$,

$$\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) \;\; = \;\; \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij})). \tag{2.2}$$

Here $\mathrm{Poi}(m)$ stands for Poisson distribution with mean m. Also, d_{ij} and $y_{i,j-1}$ are assumed to be independent.

Note that the dynamic model (2.1) produces the means and variances as

$$\begin{aligned} \mathrm{E}[Y_{ij}|\boldsymbol{x}_{ij}, z_{ij}] &= \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) = \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij})) \\ \mathrm{var}[Y_{ij}|\boldsymbol{x}_{ij}, z_{ij}] &= \sigma_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) = \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) \\ &= \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij})). \end{aligned} \tag{2.3}$$

Now, to derive the correlations between count responses under model (2.1), observe that for $j < k$, the covariance between y_{ij} and y_{ik} can be written as

$$\begin{aligned}
& \mathrm{cov}(Y_{ij}, Y_{ik} | \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \psi(z_{ij}), \psi(z_{ik})) \\
= \; & \mathrm{E}(Y_{ij} Y_{ik} | \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \psi(z_{ij}), \psi(z_{ik})) - \mathrm{E}(Y_{ij} | \boldsymbol{x}_{ij}, \psi(z_{ij})) \mathrm{E}(Y_{ik} | \boldsymbol{x}_{ik}, \psi(z_{ik})) \\
= \; & \mathrm{E}_{Y_{ij}}[Y_{ij} \mathrm{E}_{Y_{i,j+1}} \{ \ldots \mathrm{E}_{Y_{i,k-1}} (\mathrm{E}(Y_{ik} | y_{i,k-1}, y_{i,k-2}, \ldots, y_{i,j+1})) \}] \\
& - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) \mu_{ik}(\boldsymbol{\beta}, \boldsymbol{x}_{ik}, \psi(z_{ik})) \\
= \; & \sigma_{i,jk}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \psi(z_{ij}), \psi(z_{ik}), \rho) \\
= \; & \rho^{k-j} \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) \qquad (2.4)
\end{aligned}$$

(Sutradhar, 2010), yielding the correlations between y_{ij} and y_{ik} as

$$\mathrm{corr}(Y_{ij}, Y_{ik} | \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \psi(z_{ij}), \psi(z_{ik})) = \begin{cases} \rho^{k-j} \sqrt{\frac{\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}; \psi(z_{ij}))}{\mu_{ik}(\boldsymbol{\beta}, \boldsymbol{x}_{ik}; \psi(z_{ik}))}} & j < k \\ \rho^{j-k} \sqrt{\frac{\mu_{ik}(\boldsymbol{\beta}, \boldsymbol{x}_{ik}; \psi(z_{ik}))}{\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}; \psi(z_{ij}))}} & j > k. \end{cases} \qquad (2.5)$$

Notice that because the so-called error count d_{ij} in model (2.1) has a Poisson distribution with mean $\mu_{ij}(\cdot) - \rho \mu_{i,j-1}(\cdot)$, it then follows that the correlation index parameter ρ must satisfy the restriction

$$0 < \rho < \min \left[1, \frac{\mu_{ij}(\cdot)}{\mu_{i,j-1}(\cdot)} \right], \text{ for all } i = 1, \ldots, K \text{ and } j = 2, \ldots, n_i.$$

Thus, the proposed model allows only positive correlations between the repeated responses, and the exact correlation between any two responses can be computed by using (2.5).

Note that the AR(1) type dynamic model (2.1) appears to be highly practical because in practice, one expects that the correlation would decay as the time lag increases. However, such a pattern in the longitudinal setup must be influenced by the time-dependent covariates as well. The model (2.1) has also been used in time series setup especially with stationary (time independent) covariates to model the correlations of the repeated counts. See, for example, Al-Osh and Alzaid (1987), McKenzie (1988), and Winkelmann (2008) (Section 7.3). In a regression setup, for time series of counts, Zeger (1988) used correlated random effects to model the correlations of the repeated counts. This produces a complicated correlation structure without the dynamic property desired among lagged responses. If one is suspicious about the AR(1) model as opposed to any other low order auto-correlation model, then a diagnostic may be done following Sutradhar (2010) (Section 4). However, this is beyond the scope of this thesis.

There are some alternative models such as marginal models (Severini and Staniswalis, 1994, Lin and Carroll, 2001) to deal with repeated binary or count data that belong to the exponential family. These marginal models produce the mean and variance similar to (2.3). But unlike (2.5), these models do not assume any correlation structure. Consequently, for inference about these marginal models, the aforementioned authors used the so-called 'working' correlation approach. More specifically, a UNS (unstructured) correlation matrix, namely,

$$\boldsymbol{R}(=\boldsymbol{R}_i) = K^{-1}\sum_{\ell=1}^{K}\boldsymbol{r}_\ell\boldsymbol{r}_\ell^\top.$$

[see Eqn. (1.40) in Chapter 1] is used in the existing studies to construct an estimating equation for the regression parameter $\boldsymbol{\beta}$ involved in (2.3). However, the approach of using UNS in R matrix has some drawbacks. In the longitudinal semi-parametric binary fixed model setup, we demonstrate in Chapter 4 that the UNS based approach is not appropriate for this problem because it produces less efficient regression estimates than the ones obtained through the independence assumption based QL or GEE(I) approaches.

2.2 Estimation for the semi-parametric model (2.1)

Fitting of the model (2.1) to a data set requires the estimation of the nonparametric function $\psi(\cdot)$, regression parameter $\boldsymbol{\beta}$ and correlation index parameter ρ.

Because the nonparametric function $\psi(\cdot)$ is not of direct interest, Sutradhar et al. (2016) estimated $\psi(\cdot)$ with an independence assumption based SQL (semi-parametric quasi-likelihood) approach. This SQL estimate of $\psi(\cdot)$ was used to construct a SGQL (semi-parametric generalized quasi-likelihood) estimating equation (Sutradhar, 2003) for $\boldsymbol{\beta}$. The longitudinal correlation index parameter ρ was estimated using a SMM (semi-parametric method of moments) approach. For convenience, these SQL, SGQL and SMM estimating equations are presented below.

2.2.1 SQL estimation of the nonparametric function $\psi(\cdot)$

Using z_0 for z_{ij} for given i and j, Sutradhar et al. (2016) have used the independence assumption based SQL (semi-parametric quasi-likelihood) estimating equation

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\ \frac{\partial \mu_{ij}}{\partial \psi(z_0)}(\frac{y_{ij}-\mu_{ij}}{\sigma_{i,jj}}) = 0, \tag{2.6}$$

(Carota and Parmigiani, 2002) to obtain $\hat{\psi}(z_0)$, where $w_{ij}(z_0)$ is known as a kernel weight. If $w_{ij}(z_0) = 1$ for all i and j, the SQL estimating equation (2.6) reduces to the well known QL estimating equation (Wedderburn, 1974). After some algebra, it was shown in Sutradhar et al. (2016) that the SQL estimator of $\psi(\cdot)$ has the close form

$$\hat{\psi}(\boldsymbol{\beta}, z_0) \;=\; \log\left\{\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0) y_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\exp[\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta}]}\right\}. \tag{2.7}$$

2.2.2 Moment estimation for the correlation index parameter ρ

Notice from (2.4) that the lag 1 covariance has the formula

$$\mathrm{cov}(Y_{i,j-1}, Y_{ij}) = \rho\mu_{i,j-1}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j-1}, \psi(z_{i,j-1})). \tag{2.8}$$

Consequently, by equating the average sample covariance with its population counterpart in (2.8), a MM (method of moments) estimator of ρ can be obtained as

$$\begin{aligned}\hat{\rho} \;=\; & \sum_{i=1}^{K}\sum_{j=2}^{n_i}\Big[\frac{(y_{ij}-\mu_{ij}(\boldsymbol{x}_{ij}, \boldsymbol{\beta}, \psi(z_{ij})))}{\sqrt{\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))}}\Big]\Big[\frac{(y_{i,j-1}-\mu_{i,j-1}(\boldsymbol{x}_{i,j-1}, \boldsymbol{\beta}, \psi(z_{i,j-1})))}{\sqrt{\mu_{i,j-1}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j-1}, \psi(z_{i,j-1}))}}\Big] \\ & \Big/ \sum_{i=1}^{K}\sum_{j=2}^{n_i}\Big[\frac{\sqrt{\mu_{i,j-1}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j-1}, \psi(z_{i,j-1}))}}{\sqrt{\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j}, \psi(z_{i,j}))}}\Big]\end{aligned} \tag{2.9}$$

for known $\psi(\cdot)$.

However, as $\psi(\cdot)$ was estimated consistently by (2.7), the means, variances and covariances can be modified (see Sutradhar et al., 2016) as

$$\tilde{\mu}_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})) = \mathrm{E}[Y_{ij}|\boldsymbol{x}_{ij}, \hat{\psi}(\cdot)] = \exp[\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \hat{\psi}(\boldsymbol{\beta}, z_{ij})]$$

$$\tilde{\sigma}_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})) = \tilde{\mu}_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})) = \exp[\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \hat{\psi}(\boldsymbol{\beta}, z_{ij})],$$

and

$$\tilde{\sigma}_{i,jk}(\boldsymbol{\beta}, \rho, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})) = \rho^{k-j}\tilde{\mu}_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})), \text{for } j < k, \tag{2.10}$$

respectively. Next by replacing $\psi(z_{ij})$ function involved in (2.9) with $\hat{\psi}(\boldsymbol{\beta}, z_{ij})$ for known $\boldsymbol{\beta}$, we used the SMM (semi-parametric method of moments) estimator given by

$$\begin{aligned}\tilde{\rho} &= \sum_{i=1}^{K}\sum_{j=2}^{n_i}[\frac{(y_{ij}-\tilde{\mu}_{ij}(\boldsymbol{x}_{ij},\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta},z_{ij})))}{\sqrt{\tilde{\mu}_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij},\hat{\psi}(\boldsymbol{\beta},z_{ij}))}}][\frac{(y_{i,j-1}-\tilde{\mu}_{i,j-1}(\boldsymbol{x}_{i,j-1},\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta},z_{i,j-1})))}{\sqrt{\tilde{\mu}_{i,j-1}(\boldsymbol{\beta},\boldsymbol{x}_{i,j-1},\hat{\psi}(\boldsymbol{\beta},z_{i,j-1}))}}] \\ &\Bigg/\sum_{i=1}^{K}\sum_{j=2}^{n_i}[\frac{\sqrt{\tilde{\mu}_{i,j-1}(\boldsymbol{\beta},\boldsymbol{x}_{i,j-1},\hat{\psi}(\boldsymbol{\beta},z_{i,j-1}))}}{\sqrt{\tilde{\mu}_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{i,j},\hat{\psi}(z_{ij}))}}]. \qquad (2.11)\end{aligned}$$

2.2.3 Estimation of $\boldsymbol{\beta}$

As far as the estimation of the main regression parameter $\boldsymbol{\beta}$ is concerned, using

$$\begin{aligned}\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta})) &= \mathrm{E}(Y_i) = [\tilde{\mu}_{i1}(\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta})),\ldots,\ldots,\tilde{\mu}_{in_i}(\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta}))]^{\top} : n_i \times 1 \\ \tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta},\rho,\hat{\psi}(\boldsymbol{\beta})) &= \mathrm{cov}(Y_i) = (\tilde{\sigma}_{i,jk}(\boldsymbol{\beta},\rho,\hat{\psi}(\boldsymbol{\beta}))) : n_i \times n_i, \qquad (2.12)\end{aligned}$$

Sutradhar et al. (2016) have constructed the SGQL (semi-parametric generalized quasi-likelihood) estimating equation

$$\sum_{i=1}^{K}\frac{\partial[\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta}))]^{\top}}{\partial\boldsymbol{\beta}}\ [\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta},\rho,\hat{\psi}(\boldsymbol{\beta}))]^{-1}\ [\boldsymbol{y}_i-\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta},\hat{\psi}(\boldsymbol{\beta}))] = 0, \qquad (2.13)$$

for $\boldsymbol{\beta}$. This SGQL estimating equation was shown to produce both consistent and efficient estimates for $\boldsymbol{\beta}$.

Chapter 3

Semi-parametric dynamic mixed models for longitudinal count data

Panel count data analysis has been an important research topic over the last decades both in Econometrics and Statistics. See, for example, Montalvo (1997), Wooldridge (1999), Sutradhar and Bari (2007), Winkelmann (2008), Sutradhar (2011) (Chapter 8), and Sutradhar et al. (2014) for various longitudinal mixed models for such panel count data. Sutradhar and Bari (2007) have illustrated their longitudinal mixed model in the context of repeated physician office visits data. Sutradhar (2011) (Chapter 8) has also discussed an application to a panel count data consisting of repeated patent awards to selected industries in USA and their R&D (research and development) related covariates (Hausman et al., 1984, Blundell et al., 1995, Montalvo, 1997). In these panel count data models, the repeated count responses are influenced by time dependent covariates and individual common random effect. Suppose that t_{ij} denote the time at which the jth ($j = 1, \ldots, n_i$) count response is recorded from the ith ($i = 1, \ldots, K$) individual. Next, suppose that $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})^\top$ denote the $n_i \times 1$ vector of repeated count responses for the ith individual. Also suppose that the response y_{ij} of the ith individual is influenced by a fixed and known $p-$dimensional time dependent covariate vector $\boldsymbol{x}_{ij}(t_{ij})$ and another unobservable factor. We accommodate this unobservable factor by using a latent effect τ_i^*, say, which we assume to be common among the repeated counts $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$. In count data setup, these random effects are in general assumed to follow certain gamma or lognormal distributions, lognormal being most widely used (Breslow and Clayton, 1993, Schall, 1991). For this chapter, we assume that $\tau_i^* \overset{iid}{\sim} N(0, \sigma_\tau^2)$, or equivalently $\tau_i = \tau_i^*/\sigma_\tau \overset{iid}{\sim} N(0, 1)$. Note that in most of the longitudinal studies , the covariates are collected at regular intervals, $t_{ij} = hj$, where h is a constant, and hence $\boldsymbol{x}_{ij}(t_{ij})$ can be replaced by $\boldsymbol{x}_{ij}(j)$, when convenient. For example, in a physician visits study, $\boldsymbol{x}_{ij}(j)$ may represent the smoking status of the ith individual in jth month, where detailed breakdown such as smoking status over the days or weeks may not be more informative.

We now consider that y_{ij} conditional on τ_i follows a marginal Poisson distribution with mean $\mathrm{E}\left[Y_{ij}|\tau_i\right] = m_{ij}^* = \exp(\boldsymbol{x}_{ij}^\top(j)\boldsymbol{\beta} + \sigma_\tau \tau_i)$. That is,

$$m_{ij}^* = \mathrm{E}\left[Y_{ij}|\tau_i\right] = \mathrm{Var}\left[Y_{ij}|\tau_i\right] = \exp(\boldsymbol{x}_{ij}^\top(j)\boldsymbol{\beta} + \sigma_\tau \tau_i). \tag{3.1}$$

Because conditional on τ_i, $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ are likely to be correlated, some authors such as Sutradhar and Bari (2007), Sutradhar et al. (2014) modeled this through a dynamic relationship given by

$$y_{ij}|\tau_i \quad = \quad \rho * y_{i,j-1}|\tau_i + d_{ij}|\tau_i, \quad j = 2, \ldots, n_i, \tag{3.2}$$

where it is assumed that $y_{i1}|\tau_i \sim \mathrm{Poi}(m_{i1}^*)$, and for $j = 2, \ldots, n_i$, $y_{i,j-1}|\tau_i \sim \mathrm{Poi}(m_{i,j-1}^*)$, and $d_{ij}|\tau_i \sim \mathrm{Poi}(m_{ij}^* - \rho m_{i,j-1}^*)$. Here $\mathrm{Poi}(m)$ stands for Poisson distribution with mean m. In (3.2), conditional on τ_i, d_{ij} and $y_{i,j-1}$ are independent. Furthermore, for a given count $y_{i,j-1}$, $\rho * y_{i,j-1} = \sum_{k=1}^{y_{i,j-1}} b_k(\rho)$ is a binomial thinning operation, where $b_k(\rho)$ stands for a Bernoulli random variable with $\Pr[b_k(\rho) = 1] = \rho$ and $\Pr[b_k(\rho) = 0] = 1-\rho$. This model (3.2) produces the pairwise correlations conditional on τ_i as

$$\mathrm{Corr}(Y_{ij}, Y_{ik}|\tau_i) = \rho^{k-j}\sqrt{m^*_{ij}/m^*_{ik}} \quad \text{for } j < k.$$

In some practical situations, it may happen that in addition to $\boldsymbol{x}_{ij}(t_{ij})$ or $\boldsymbol{x}_{ij}(j)$, some other secondary covariates are collected from the ith individual at times t_{ij} $(j = 1, \cdots, n_i)$. We consider a scalar secondary covariate for convenience, namely, $z_{ij}(t_{ij})$. For example, in the Health Care Utilization (HCU) data (Sutradhar, 2003), the repeated numbers of yearly physician visits were studied as a function of various covariates such as gender, education level, chronic disease status and age of the individuals. Here the effects of gender, education level and chronic disease status on individuals' yearly physician office visits may be of primary interest, whereas the age of an individual could be considered as the secondary covariate z_{ij}. On top of $\boldsymbol{x}_{ij}(t_{ij})$, this secondary covariate $z_{ij}(t_{ij})$ must influence y_{ij} as well. But its effect is not of direct interest. Following some semi-parametric longitudinal fixed models (Severini and Staniswalis, 1994, Zeger and Diggle, 1994, Sneddon and Sutradhar, 2004, Lin and Carroll, 2001, 2006, You and Chen, 2007, Warriyar and Sutradhar, 2014), we accommodate the effect of $z_{ij}(t_{ij})$ nonparametrically. The ultimate model for the responses $y_{ij}, j = 1, \cdots, n_i$, as a function of the fixed covariate $\boldsymbol{x}_{ij}(t_{ij})$, random effect τ_i and secondary covariate $z_{ij}(t_{ij})$ will be referred to as the semi-parametric generalized linear mixed model (SGLMM) for longitudinal data. This mixed model is discussed in details in the next section. Note that this SGLMM for longitudinal data is new, and it has not been adequately addressed in the literature.

For the purpose of fitting the SGLMM to a longitudinal data set, a step by step estimation for the nonparametric function, and regression, overdispersion, and longitudinal correlation index parameters, is given in Section 3.2. The asymptotic properties of the estimators are discussed in Section 3.3 in details. In Section 3.4, we carry out an extensive simulation study to examine the finite sample performance of the estimation approaches for the proposed semi-parametric dynamic mixed model.

3.1 Proposed SGLMM for longitudinal count data and its basic properties

In this section, we extend the GLMM for longitudinal data ((3.1)−(3.2)) for count data to the semi-parametric setup. For the purpose, we add a nonparametric function $\psi(z_{ij})$ to the linear predictor $\boldsymbol{x}_{ij}^{\top}(j)\boldsymbol{\beta} + \sigma_\tau \tau_i$ in the mixed model (3.1). One may write the conditional marginal mean and variance as

$$\mu_{ij}^{*} = \mathrm{E}\left[Y_{ij}|\tau_i\right] = \mathrm{Var}\left[Y_{ij}|\tau_i\right] = \exp\left\{\boldsymbol{x}_{ij}^{\top}(j)\boldsymbol{\beta} + \sigma_\tau \tau_i + \psi(z_{ij})\right\}, \tag{3.3}$$

which are now semi-parametric because of the introduction of $\psi(z_{ij})$. This model (3.3) is known as SGLMM for longitudinal Poisson count data. However, to be brief, we may refer to this SGLMM (3.3) as the semi-parametric mixed model (SMM). Note that τ_i in (3.3), similar to the mixed model (3.1)−(3.2), will be assumed to follow $\tau_i \overset{iid}{\sim} N(0,1)$ (see also Breslow and Clayton, 1993, Schall, 1991).

The correlations among $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ conditional on τ_i are modeled through a dynamic relationship given by

$$y_{ij}|\tau_i \;=\; \rho * y_{i,j-1}|\tau_i + d_{ij}|\tau_i\,, \tag{3.4}$$

which is similar to but different from (3.2). The difference lies in the fact that the conditional marginal means and variances under (3.4) have the form (3.3), whereas under the mixed model (3.1) and (3.2), $y_{ij}|\tau_i \sim \text{Poi}(m^*_{ij})$ with $m^*_{ij} = \exp(\boldsymbol{x}^\top_{ij}(j)\boldsymbol{\beta}+\sigma\tau_i)$. Consequently, following the correlation properties under the model (3.1)$-$(3.2), we can write the formula for correlations under the present model (3.3)$-$(3.4) as

$$\begin{aligned}
\text{Corr}(Y_{ij}, Y_{ik}|\tau_i) &= \rho^{k-j}\sqrt{\mu^*_{ij}/\mu^*_{ik}} \qquad \text{for } j < k, \text{ or} \\
\text{Cov}(Y_{ij}, Y_{ik}|\tau_i) &= \rho^{k-j}\mu^*_{ij} \qquad \text{for } j < k.
\end{aligned} \tag{3.5}$$

Next, we provide the basic properties of the count responses, i.e., the unconditional mean, variance and pairwise covariances under the proposed SMM ((3.3)$-$(3.4)) as in the following lemma.

Lemma 3.1. *Under the SMM (3.3)$-$(3.4), the responses have the following moment properties:*

$$\mu_{ij} \equiv \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau, \psi(\cdot)) = \text{E}\left[Y_{ij}\right] = \exp\left\{\boldsymbol{x}^\top_{ij}\boldsymbol{\beta} + \frac{\sigma^2_\tau}{2} + \psi(z_{ij})\right\}, \tag{3.6}$$

$$\sigma_{ijj} \equiv \sigma_{ijj}(\boldsymbol{\beta}, \sigma_\tau, \psi(\cdot)) = \text{Var}\left[Y_{ij}\right] = \mu_{ij} + \mu^2_{ij}\left[\exp(\sigma^2_\tau) - 1\right], \text{ and} \tag{3.7}$$

$$\begin{aligned}
\sigma_{ijk} &\equiv \sigma_{ijk}(\boldsymbol{\beta}, \sigma_\tau, \rho, \psi(z_0)) = \text{Cov}\left(Y_{ij}, Y_{ik}\right) \\
&= \rho^{k-j}\mu_{ij} + \mu_{ij}\mu_{ik}\left(\exp(\sigma^2_\tau) - 1\right), \quad j < k, \quad j, k = 1, \cdots, n_i\,.
\end{aligned} \tag{3.8}$$

Proof. The proof follows from the moment generating function of the normal distribution.

$$\begin{aligned}
\mu_{ij} &= \text{E}\left[Y_{ij}\right] = \text{E}\left[\text{E}\left(Y_{ij}|\tau_i\right)\right] \\
&= \exp\left\{\boldsymbol{x}^\top_{ij}\boldsymbol{\beta} + \psi(z_{ij})\right\}\text{E}\left(\exp(\sigma_\tau\tau_i)\right) = \exp\left\{\boldsymbol{x}^\top_{ij}\boldsymbol{\beta} + \frac{\sigma^2_\tau}{2} + \psi(z_{ij})\right\},
\end{aligned}$$

and

$$\begin{aligned}
\sigma_{ijj} &= \mathrm{Var}\left[Y_{ij}\right] = \mathrm{Var}\left[\mathrm{E}\left(Y_{ij}|\tau_i\right)\right] + \mathrm{E}\left[\mathrm{Var}\left(Y_{ij}|\tau_i\right)\right] \\
&= \mathrm{Var}\left[\exp\left\{\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \psi(z_{ij}) + \sigma_\tau\tau_i\right\}\right] + \mathrm{E}\left[\mu_{ij}^*\right] \\
&= \exp\left\{2\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + 2\psi(z_{ij})\right\}\mathrm{Var}\left(\exp(\sigma_\tau\tau_i)\right) + \mu_{ij} \\
&= \left[\exp\left\{\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \psi(z_{ij}) + \frac{\sigma_\tau^2}{2}\right\}\right]^2\left(\exp(\sigma_\tau^2) - 1\right) + \mu_{ij} \\
&= \mu_{ij} + \mu_{ij}^2\left[\exp(\sigma_\tau^2) - 1\right].
\end{aligned}$$

Similarly, the formula for the pair-wise covariances in (3.8) is derived as

$$\sigma_{ijk} = \mathrm{Cov}\left(Y_{ij}, Y_{ik}\right) = \mathrm{E}\left[\mathrm{Cov}\left(Y_{ij}, Y_{ik}|\tau_i\right)\right] + \mathrm{Cov}\left[\mathrm{E}\left(Y_{ij}|\tau_i\right), \mathrm{E}\left(Y_{ik}|\tau_i\right)\right],$$

which by using (3.5) reduces to

$$\begin{aligned}
\sigma_{ijk} &= \mathrm{E}\left[\rho^{k-j}\mu_{ij}^*\right] + \mathrm{Cov}\left[\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \psi(z_{ij}) + \sigma_\tau\tau_i), \exp(\boldsymbol{x}_{ik}^\top\boldsymbol{\beta} + \psi(z_{ik}) + \sigma_\tau\tau_i)\right] \\
&= \rho^{k-j}\mu_{ij} + \mu_{ij}\mu_{ik}\left(\exp(\sigma_\tau^2) - 1\right).
\end{aligned}$$

$\square$

Note that the above notations, i.e., $\mu_{ij}(\cdot)$, $\sigma_{ijj}(\cdot)$ and $\sigma_{ijk}(\cdot)$ were also used in Chapter 2, specifically in the equations (2.3)−(2.4) [see also (1.36)]. However, they were written under the fixed model, i.e., for the cases where $\sigma_\tau = 0$.

We remark that the Lemma 3.1 further gives the lag $(k-j)$ unconditional correlations as

$$\mathrm{Corr}\left(Y_{ij}, Y_{ik}\right) = \frac{\sigma_{ijk}}{\sqrt{\sigma_{ijj}\sigma_{ikk}}}$$

$$= \frac{\mu_{ij}\rho^{k-j} + \mu_{ij}\mu_{ik}\left(\exp(\sigma_\tau^2) - 1\right)}{\left[\left\{\mu_{ij} + \mu_{ij}^2\left(\exp(\sigma_\tau^2) - 1\right)\right\}\left\{\mu_{ik} + \mu_{ik}^2\left(\exp(\sigma_\tau^2) - 1\right)\right\}\right]^{\frac{1}{2}}}. \tag{3.9}$$

We further remark that because in model (3.4) $d_{ij}|\tau_i \sim \text{Poi}(\mu_{ij}^* - \rho\mu_{i,j-1}^*)$ with $(\mu_{ij}^* - \rho\mu_{i,j-1}^*) \geq 0$, the correlation index parameter ρ in (3.8) or (3.9), must satisfy the range restriction $0 \leq \rho < \min[1, \mu_{ij}^*/\mu_{i,j-1}^*]$, which is the same as

$$0 \leq \rho < \min[1, \nu_{ij}^*/\nu_{i,j-1}^*] \text{ for } j = 2, \ldots, n_i \text{ and } i = 1, \ldots, K, \tag{3.10}$$

where $\nu_{ij}^* = \exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \psi(z_{ij}))$. Because consecutive observations tend to have similar covariate values, $\nu_{ij}^*/\nu_{i,j-1}^*$ in (3.10) are likely to be close to 1, so practically (3.10) should not force a strict restriction on data.

Notice from (3.9) that under the proposed model, unlike the longitudinal fixed model case (2.5), the correlation index parameter value $\rho = 0$ does not imply that the responses under the present model (3.3)$-$(3.4) are uncorrelated. The repeated responses are uncorrelated only when both $\rho = 0$ and $\sigma_\tau^2 = 0$. However, since in the mixed model $\sigma_\tau^2 > 0$ always, the pairwise responses are positively correlated irrespective of the case whether ρ is zero or not. This correlation behavior of the proposed model will be exploited in the next section in order to develop the necessary estimating equations.

3.2 Quasi-likelihood estimation for the proposed SGLMM for longitudinal count data

In (3.9) we have seen that the repeated responses are uncorrelated only when both correlation index parameter ρ and over-dispersion effect parameter σ_τ are zero. Therefore, one has to be careful while estimating the regression effects $\boldsymbol{\beta}$ and nonparametric function $\psi(z_{ij}(t_{ij}))$ using any GEE(I) (the generalized estimating equation based on the assumption of independence) approach. In fact, although one can argue that it would be okay to use $\rho = 0$ for initial estimation of these parameters and functions, one cannot set $\sigma_\tau^2 = 0$ because this would always produce inconsistent estimates since σ_τ^2 is involved in the mean function (3.6) along with them [$\boldsymbol{\beta}$ and $\psi(z_{ij})$]. As opposed to the semi-parametric longitudinal fixed model (Severini and Staniswalis, 1994, Lin and Carroll, 2001, 2006, You and Chen, 2007, Warriyar and Sutradhar, 2014) this is a major additional estimation problem in the present semi-parametric longitudinal mixed model case.

In this section we develop a quasi-likelihood estimation approach which provides consistent estimates for all parameters and the nonparametric function involved in the SGLMM. Note that this approach has been used by some authors (Severini and Staniswalis, 1994, Lin and Carroll, 2001, 2006, Warriyar and Sutradhar, 2014) for the SGLFM (semi-parametric generalized linear fixed model) for longitudinal data. Severini and Staniswalis (1994) and Lin and Carroll (2001) (see also Zeger and Diggle, 1994) refer to their procedure as the semi-parametric generalized estimating equation (SGEE) approach which does not need any specification of the underlying longitudinal correlation structure. However, there has been many studies showing that, under specific circumstances, independence assumption based GEE (GEE(I)) approach may produce more efficient regression estimates at times than arbitrary 'working' correlations based GEE approach. See for example, Sutradhar (2010) (Section 3.1) (see also Sutradhar and Das, 1999) in the context of GLFM for longitudinal count data. Also, as we will show in Chapter 4, in the context of semi-parametric longitudinal models for binary data, the SGEE(I) approach produces more efficient regression estimates as compared to 'working' SGEE approaches. This efficiency feature undermines the use of the GEE or SGEE approaches. Thus, we do not discuss the GEE approaches any further in this chapter. Instead, we assume that the repeated count data are generated following the AR(1) Poisson mixed model (3.4) based correlation structure (3.9) and consequently use the true correlation structure based semi-parametric GQL (generalized quasi-likelihood) approach for the estimation of the main regression effects (of the primary covariates) and the overdispersion parameter (Sutradhar and Bari (2007); Sutradhar (2011, Chapter 8)). Next, because the nonparametric function and the longitudinal correlations are of secondary interest, we estimate them using the simpler SQL (semi-parametric QL) and SMM (semi-parametric method of moments) approaches, respectively, as opposed to the SGQL approach. These estimation approaches are discussed in the following subsections.

3.2.1 QL estimation for the nonparametric function $\psi(\cdot)$

The function $\psi(z_{ij})$ has to be estimated for all $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$, where z_{ij} is a secondary covariate collected at time t_{ij}. Thus, it is equivalent to estimate $\psi(z_0)$, say, where $z_0 \equiv z_{ij}$ for all values of i and j. In the SGLFM setup for longitudinal data, some authors such as Lin and Carroll (2001) (see also Severini and Staniswalis, 1994) have estimated the function $\psi(\cdot)$ by using a 'working' correlation structure based estimating equation approach. There are several drawbacks of this GEE approach. For example, these authors have considered the case $n_i = n$, say, for estimating their so-called $n \times n$ unstructured 'working' common correlation matrix, whereas in practice n_i's can be different. Furthermore, using a common average 'working' correlation matrix for all individuals may not be appropriate mainly because actual correlations may be functions of time dependent covariates [see (3.9)]. These limitations restrict the application of their approach to longitudinal problems. Our study in this thesis indicates that because z_{ij} are simply fixed covariates, for consistent estimation of $\psi(z_{ij})$, which is of secondary interest, it would be enough to use an independent assumption based estimating equation, whereas the main regression parameter (effect of primary covariates) $\boldsymbol{\beta}$ would be estimated consistently and as efficiently as possible by using the correlation structure based estimating equation. Furthermore unlike the existing fixed regression models, we also need to consistently and efficiently estimate the overdispersion parameter, σ_τ^2, involved in the present mixed model (3.3).

In the quasilikelihood (QL) approach for independent data (Wedderburn, 1974) one explores the mean and the variance functions, variance being a function of mean such as in a GLM setup, to write a QL estimating equation for the parameter involved in the mean function. When the mean function involves a nonparametric function, one way to address the estimation of such a function is by solving a kernel weights based semi-parametric QL (SQL) estimating equation. For the estimation of $\psi(z_0)$ in the present setup which influences the mean function $\mu_{ij}(\boldsymbol{\beta}, \sigma_\tau, \psi(z_0))$, the SQL estimating equation has the form

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\frac{\partial \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau, \psi(z_0))}{\partial \psi(z_0)}\left(\frac{y_{ij} - \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau, \psi(z_0))}{\sigma_{ijj}(\boldsymbol{\beta}, \sigma_\tau, \psi(z_0))}\right) = 0 \qquad (3.11)$$

(e.g. Carota and Parmigiani, 2002, Sutradhar et al., 2016, see also (2.6)), where $w_{ij}(z_0)$ is referred to as the kernel weight defined as

$$w_{ij}(z_0) = p_{ij}(\frac{z_0 - z_{ij}}{b}) \Big/ \sum_{l=1}^{K}\sum_{u=1}^{n_l} p_{lu}(\frac{z_0 - z_{lu}}{b}) \qquad (3.12)$$

where p_{ij} is the kernel density with a suitable bandwidth parameter, b. Note that this SQL estimating equation (3.11) is different than the so-called 'working' correlations based SGEE (semi-parametric GEE) used by Lin and Carroll (2001, 2006) (see also Severini and Staniswalis, 1994). It is simpler than SGEE and also it assures the consistency of the estimator. Notice that even though SGEE is developed for efficient estimation, it may produce inefficient estimate than the 'working' independence based SQL estimator [Lin and Carroll (2001, Section 7), Sutradhar et al. (2016)].

With regard to the selection of the kernel density $p_{ij}(\cdot)$, it should be noted that there is, in fact, no unique choice for the selection of such a density. Some of the widely used kernel densities, for example, are the Gaussian density given by

$$p_{ij}\left(\frac{z_0 - z_{ij}}{b}\right) = \frac{1}{\sqrt{2\pi}\, b} \exp\left\{-\frac{1}{2}\left(\frac{z_0 - z_{ij}}{b}\right)^2\right\}, \tag{3.13}$$

and the Epanechnikov kernel (Pagan and Ullah (1999, p. 28)) with density

$$p_{ij}(\varphi) = \begin{cases} \frac{1}{4}(1-\varphi^2) & \text{for } |\varphi| \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \varphi = \frac{z_0 - z_{ij}}{b}. \tag{3.14}$$

In (3.12)−(3.14), b is a suitable bandwidth parameter. We assume that this parameter is chosen such that the mean squared error of the estimator of $\psi(z_{ij})$ will be minimized. It has been suggested to choose b as $b \propto K^{-1/5}$ (Altman, 1990, Powell and Stoker, 1996, Pagan and Ullah, 1999, Horowitz, 2009).

Now, because $\partial\mu_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_0))/\partial\psi(z_0) = \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_0))$, the SQL estimating equation (3.11) can be further simplified as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left(\frac{y_{ij} - \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_0))}{1 + \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_0))\left(\exp(\sigma_\tau^2) - 1\right)}\right) = 0, \tag{3.15}$$

which, for given values of $\boldsymbol{\beta}$ and σ_τ^2, may be solved iteratively until convergence.

Notice that the estimate of $\psi(z_0)$ from the SQL estimating equation (3.15) is a function of $\boldsymbol{\beta}$ and σ_τ^2. Hence we denote the estimator of $\psi(z_0)$ by $\hat{\psi}(z_0; \boldsymbol{\beta}, \sigma_\tau^2)$. The consistency property of this estimator is discussed in Section 3.3, and we study through simulations its finite sample properties along with the properties of other estimators in Section 3.4.

3.2.2 SGQL estimation of regression effects $\boldsymbol{\beta}$

Recall that the means $\mu_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ij}))$ and variances $\sigma_{ijj}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ij}))$ for all $j = 1, \cdots, n_i; i = 1, \cdots, K$ are used to estimate $\psi(\cdot)$ by using the well-known QL approach (Wedderburn, 1974). As mentioned, this was done to obtain consistent estimator of $\psi(\cdot)$. In this section, however, we develop the estimation technique to obtain both consistent and efficient estimates for the regression parameters. For this purpose we need to consider the covariances among the repeated responses. Thus, we now construct the mean vector and covariance matrix of the repeated count responses. For known $\psi(\cdot)$, let

$$\begin{aligned} & \mathrm{E}(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \sigma_\tau^2, \psi(\cdot)) \\ = \; & (\mu_{i1}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(\cdot)), \ldots, \mu_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(\cdot)), \ldots, \mu_{in_i}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(\cdot)))^\top, \end{aligned} \tag{3.16}$$

and

$$\mathrm{Cov}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \psi(\cdot)) = (\sigma_{ijk}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \psi(\cdot))) : \quad n_i \times n_i\,, \tag{3.17}$$

where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})^\top$ is the $n_i \times 1$ vector of responses for the ith individual. However, it is clear from the last section that when $\psi(z_{ij})$ are estimated by solving the SQL estimating equation (3.15), we obtain the estimator $\hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2)$ which contains unknown $\boldsymbol{\beta}$ and σ_τ^2. Consequently, the mean vector and the covariance matrix now have the forms

$$\begin{aligned} \bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)) \; = \; & \Big(\bar{\mu}_{i1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)), \ldots, \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)), \ldots, \\ & \bar{\mu}_{in_i}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))\Big)^\top : \quad n_i \times 1 \qquad \text{and} \end{aligned} \tag{3.18}$$

$$\bar{\mathbf{\Sigma}}_i(\boldsymbol{\beta}, \sigma^2_\tau, \rho, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau)) \;\; = \;\; (\bar{\sigma}_{ijk}(\boldsymbol{\beta}, \sigma^2_\tau, \rho, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau))) : \quad n_i \times n_i. \tag{3.19}$$

We now use these new notations from (3.18) and (3.19) and following Sutradhar (2003), for example, construct the semi-parametric GQL (SGQL) estimating equation for $\boldsymbol{\beta}$ as

$$\begin{aligned} & \sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau))}{\partial \boldsymbol{\beta}} \bar{\mathbf{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma^2_\tau, \rho, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau)) \\ \times & \Big(\boldsymbol{y}_i - \bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau))\Big) = 0. \end{aligned} \tag{3.20}$$

Note that the computation of the derivative matrix $\frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau))}{\partial \boldsymbol{\beta}}$ in (3.20) requires the formula for the derivative $\frac{\partial \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau)}{\partial \boldsymbol{\beta}}$, whereas this derivative would have been zero if $\psi(\cdot)$ was known. The exact formula for the gradient matrix $\frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau))}{\partial \boldsymbol{\beta}}$ may be computed as

$$\begin{aligned} & \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau))}{\partial \boldsymbol{\beta}} \\ = & \frac{\partial(\bar{\mu}_{i1}(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau)), \ldots, \bar{\mu}_{in_i}(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(\boldsymbol{\beta}, \sigma^2_\tau)))}{\partial \boldsymbol{\beta}}, \end{aligned} \tag{3.21}$$

where, for $j = 1, \ldots, n_i$, one obtains

$$\begin{aligned} & \frac{\partial \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma^2_\tau))}{\partial \boldsymbol{\beta}} \\ = & \; \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma^2_\tau, \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma^2_\tau)) \left[\boldsymbol{x}_{ij} + \frac{\partial \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma^2_\tau)}{\partial \boldsymbol{\beta}}\right]. \end{aligned} \tag{3.22}$$

Now to compute the derivative $\frac{\partial \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma^2_\tau)}{\partial \boldsymbol{\beta}}$ for (3.22), we turn back to the estimating equation (3.15) for $\psi(z_0)$, and take its derivative with respect to $\boldsymbol{\beta}$ and obtain

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{\frac{(\exp(\sigma_\tau^2)-1)\,y_{ij}+1}{\left[1+\bar{\mu}_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2))\,(\exp(\sigma_\tau^2)-1)\right]^2}\right\}$$
$$\times \quad \hat{\mu}_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2))\left[\boldsymbol{x}_{ij}+\frac{\partial\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2)}{\partial\boldsymbol{\beta}}\right]=0,$$

yielding

$$\frac{\partial\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2)}{\partial\boldsymbol{\beta}} \tag{3.23}$$
$$= \quad \frac{-\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{\frac{1+y_{ij}\left(\exp(\sigma_\tau^2)-1\right)}{\left[1+\bar{\mu}_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2))(\exp(\sigma_\tau^2)-1)\right]^2}\right\}\bar{\mu}_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2))\,\boldsymbol{x}_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{\frac{1+y_{ij}(\exp(\sigma_\tau^2)-1)}{\left[1+\bar{\mu}_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2))(\exp(\sigma_\tau^2)-1)\right]^2}\right\}\bar{\mu}_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2))}.$$

Applying the gradient function from (3.22) to (3.20), we now solve Eqn. (3.20). The SGQL estimate of $\boldsymbol{\beta}$ obtained by solving (3.20) will be denoted by $\hat{\boldsymbol{\beta}}$. Its asymptotic and finite sample properties are discussed in Sections 3.3 and 3.4, respectively.

3.2.3 SGQL estimation of the random effect variance σ_τ^2

In general, the generalized method of moments (GMM) and the generalized quasi-likelihood (GQL) are popular procedures for the estimation of the overdispersion index parameter, σ_τ^2, involved in the SGLMM (3.2) with repeated count data. However, it was demonstrate by Rao et al. (2012) (see also Sutradhar (2011, Chapter 8, table 8.2)), under a linear longitudinal setup, that the GQL approach produces more efficient estimate for this parameter as compared to the GMM approach. Furthermore, Sutradhar and Bari (2007) demonstrated that, for count data, the GQL approach also performs well in estimating this parameter under a longitudinal setup.

In this section, we generalize the GQL approach to the semi-parametric longitudinal setup. Note that our proposed estimating equation would be similar to (3.20) for $\boldsymbol{\beta}$ estimation. The difference lies in the fact that the SGQL estimating equation for σ_τ^2 will be constructed using second order responses.

3.2.3.1 SGQL estimation using squared responses

Consider a vector of squared responses $\boldsymbol{U}_i = [Y_{i1}^2, \ldots, Y_{ij}^2, \ldots, Y_{in_i}^2]^\top$. Then a GQL estimating equation for σ_τ^2 may be developed by minimizing the quadratic distance function

$$Q = (\boldsymbol{u}_i - E[\boldsymbol{U}_i])^\top \{\mathrm{Cov}[\boldsymbol{U}_i]\}^{-1} (\boldsymbol{u}_i - E[\boldsymbol{U}_i]) \tag{3.24}$$

(Sutradhar and Bari, 2007), where $\boldsymbol{u}_i$ is the observed value of $\boldsymbol{U}_i$. For the computation of $E[\boldsymbol{U}_i]$ and $\mathrm{Cov}[\boldsymbol{U}_i]$ in the present semi-parametric setup, we first recall from (3.18) that $\mu_{ij} = E[Y_{ij}]$ now has the formula $\bar{\mu}_{ij} \equiv \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))$. This is because the estimate of $\psi(z_0)$, $\hat{\psi}(z_0; \boldsymbol{\beta}, \sigma_\tau^2) \equiv \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)$ (3.15), is still a function of unknown $\boldsymbol{\beta}$ and σ_τ^2. Then, we may compute $E[\boldsymbol{U}_i]$ by using

$$E[Y_{ij}^2] = \bar{\lambda}_{ijj}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot)) = \bar{\mu}_{ij} + \bar{\mu}_{ij}^2\, e^{\sigma_\tau^2}. \tag{3.25}$$

More specifically,

$$\begin{aligned} \bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \psi(\cdot)) &= \mathrm{E}(\boldsymbol{U}_i) = E[Y_{i1}^2, \ldots, Y_{ij}^2, \ldots, Y_{in_i}^2]^\top \\ &= [\bar{\lambda}_{i11}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot)), \ldots, \bar{\lambda}_{ijj}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot)), \ldots, \bar{\lambda}_{in_in_i}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))]^\top. \end{aligned} \tag{3.26}$$

By using similar notation we now compute $\bar{\boldsymbol{\Omega}}_i = \mathrm{Cov}(\boldsymbol{U}_i)$. To be brief, we use $\bar{\mu}_{ij}$ for $\bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))$, and $\bar{\lambda}_{ijj}$ for $\bar{\lambda}_{ijj}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))$. The diagonal elements of $\bar{\boldsymbol{\Omega}}_i$ can be obtained following Sutradhar and Bari (2007, Section 3) as

$$\mathrm{Var}\left(Y_{ij}^2\right) = \bar{\mu}_{ij}\left[1 + 7\bar{\mu}_{ij}\,\exp(\sigma_\tau^2) + 6\bar{\mu}_{ij}^2\,\exp(3\sigma_\tau^2) + \bar{\mu}_{ij}^3\,\exp(6\sigma_\tau^2)\right] - \bar{\lambda}_{ijj}^2. \tag{3.27}$$

3.2.3.1.1 Computation of the off-diagonal elements

To compute the off-diagonal elements of $\bar{\boldsymbol{\Omega}}_i$, we will use the following 3 lemmas. First, model (3.4) leads directly to the recursive conditional expectation formula

$$\mathrm{E}\left[\left(Y_{ij} - \mu_{ij}^*\right)\middle|\, Y_{i,j-1} = y_{i,j-1}, \tau_i\right] = \rho\left(y_{i,j-1} - \mu_{i,j-1}^*\right), \tag{3.28}$$

as well as the following conditional expectation formula:

Lemma 3.2. *Lag 1 expectation of conditional corrected squares: for* $j = 2, \cdots, n_i$,

$$\begin{aligned}\mathrm{E}\left[\left\{\left(Y_{ij} - \mu_{ij}^*\right)^2 - \mu_{ij}^*\right\}\middle|\, Y_{i,j-1} = y_{i,j-1}, \tau_i\right] &= \rho^2\left[\left(y_{i,j-1} - \mu_{i,j-1}^*\right)^2 - \mu_{i,j-1}^*\right] \\ &\quad + \rho\left(1-\rho\right)\left(y_{i,j-1} - \mu_{i,j-1}^*\right).\end{aligned} \tag{3.29}$$

Proof. By expanding the left hand side of (3.29), we obtain

$$\begin{aligned}&\mathrm{E}\left[\left(Y_{ij} - \mu_{ij}^*\right)^2\middle|\, Y_{i,j-1} = y_{i,j-1}, \tau_i\right] = \mathrm{E}\left[\left(Y_{ij}^2 - 2\mu_{ij}^* Y_{ij} + {\mu_{ij}^*}^2\right)\middle|\, Y_{i,j-1} = y_{i,j-1}, \tau_i\right] \\ &= \rho^2 y_{i,j-1}^2 - 2\rho^2\mu_{i,j-1}^* y_{i,j-1} - \rho^2 y_{i,j-1} + \rho\, y_{i,j-1} + \rho^2{\mu_{i,j-1}^*}^2 - \rho\,\mu_{i,j-1}^* + \mu_{ij}^* \\ &= \rho^2\left(y_{i,j-1} - \mu_{i,j-1}^*\right)^2 + \rho\left(y_{i,j-1} - \mu_{i,j-1}^*\right) - \rho^2\left(y_{i,j-1} - \mu_{i,j-1}^*\right) - \rho^2\mu_{i,j-1}^* + \mu_{ij}^* \\ &= \rho^2\left[\left(y_{i,j-1} - \mu_{i,j-1}^*\right)^2 - \mu_{i,j-1}^*\right] + \rho\left(1-\rho\right)\left(y_{i,j-1} - \mu_{i,j-1}^*\right) + \mu_{ij}^*,\end{aligned}$$

yielding the lemma. □

Lemma 3.3. *Lag* $(k-j)$ *expectation of conditional corrected squares: for* $j < k$ *and*

$j, k = 1, \cdots, n_i$,

$$\mathrm{E}\left[\left\{(Y_{ik} - \mu^*_{ik})^2 - \mu^*_{ik}\right\} \middle| Y_{i,j} = y_{i,j}, \tau_i\right] = \rho^{2(k-j)}\left[\left(y_{ij} - \mu^*_{ij}\right)^2 - \mu^*_{ij}\right] + \rho^{k-j}\left(1 - \rho^{k-j}\right)\left(y_{ij} - \mu^*_{ij}\right). \tag{3.30}$$

Proof. It follows from (3.29) that conditional on τ_i,

$$\begin{aligned}
&\mathrm{E}\left[\left\{(Y_{ik} - \mu^*_{ik})^2 - \mu^*_{ik}\right\} \middle| Y_{i,j} = y_{i,j}, \tau_i\right] \\
&= \mathrm{E}\left[\rho^2\left\{\rho^2\left[\left(y_{i,k-2} - \mu^*_{i,k-2}\right)^2 - \mu^*_{i,k-2}\right] + \rho(1-\rho)\left(y_{i,k-2} - \mu^*_{i,k-2}\right)\right\}\right. \\
&\quad \left. + \rho(1-\rho)\rho\left(y_{i,k-2} - \mu^*_{i,k-2}\right) \middle| Y_{i,j} = y_{i,j}, \tau_i\right] \\
&= \mathrm{E}\left[\rho^4\left[\left(y_{i,k-2} - \mu^*_{i,k-2}\right)^2 - \mu^*_{i,k-2}\right] + \rho(1-\rho)\left(\rho + \rho^2\right)\left(y_{i,k-2} - \mu^*_{i,k-2}\right) \middle| Y_{i,j} = y_{i,j}, \tau_i\right].
\end{aligned}$$

Further, using (3.28) and (3.29) recursively, we obtain

$$\begin{aligned}
&\mathrm{E}\left[\left\{(Y_{ik} - \mu^*_{ik})^2 - \mu^*_{ik}\right\} \middle| Y_{i,j} = y_{i,j}, \tau_i\right] \\
&= \rho^{2(k-j)}\left[\left(y_{ij} - \mu^*_{ij}\right)^2 - \mu^*_{ij}\right] + \rho(1-\rho)\left(\rho^{k-j-1} + \ldots + \rho^{2(k-j-1)}\right)\left(y_{ij} - \mu^*_{ij}\right) \\
&= \rho^{2(k-j)}\left[\left(y_{ij} - \mu^*_{ij}\right)^2 - \mu^*_{ij}\right] + \rho^{k-j}(1-\rho)\left(1 + \ldots + \rho^{k-j-1}\right)\left(y_{ij} - \mu^*_{ij}\right) \\
&= \rho^{2(k-j)}\left[\left(y_{ij} - \mu^*_{ij}\right)^2 - \mu^*_{ij}\right] + \rho^{k-j}\left(1 + \ldots + \rho^{k-j-1} - \rho - \ldots - \rho^{k-j}\right)\left(y_{ij} - \mu^*_{ij}\right) \\
&= \rho^{2(k-j)}\left[\left(y_{ij} - \mu^*_{ij}\right)^2 - \mu^*_{ij}\right] + \rho^{k-j}\left(1 - \rho^{k-j}\right)\left(y_{ij} - \mu^*_{ij}\right).
\end{aligned}$$

□

Lemma 3.4. *Unconditional product moments: For $j < k$ $(j, k = 1, \cdots, n_i)$,*

$$\mathrm{E}(Y_{ij}^2 Y_{ik}^2) = 2\rho^{2(k-j)}\mu_{ij}^2\, e^{\sigma_\tau^2} + 4\rho^{k-j}\,\mu_{ik}\mu_{ij}^2\, e^{3\sigma_\tau^2} + 2\rho^{k-j}\,\mu_{ij}^2\, e^{\sigma_\tau^2} + 2\rho^{k-j}\,\mu_{ik}\mu_{ij}\, e^{\sigma_\tau^2} + \rho^{k-j}\,\mu_{ij} + \mu_{ik}^2\mu_{ij}^2\, e^{6\sigma_\tau^2} + \mu_{ik}\mu_{ij}^2\, e^{3\sigma_\tau^2} + \mu_{ik}^2\mu_{ij}\, e^{3\sigma_\tau^2} + \mu_{ik}\mu_{ij}\, e^{\sigma_\tau^2}. \tag{3.31}$$

Proof. By (3.28) and Lemma 3.3, we compute

$$
\begin{aligned}
&\mathrm{E}(Y_{ij}^2 Y_{ik}^2|\tau_i) = \mathrm{E}\left[\left(Y_{ij} - \mu_{ij}^* + \mu_{ij}^*\right)^2 \left(Y_{ik} - \mu_{ik}^* + \mu_{ik}^*\right)^2 |\tau_i\right] \\
&= \mathrm{E}\left[\left\{\left(Y_{ij} - \mu_{ij}^*\right)^2 + 2\mu_{ij}^* \left(Y_{ij} - \mu_{ij}^*\right) + {\mu_{ij}^*}^2\right\} \left\{\left(Y_{ik} - \mu_{ik}^*\right)^2 + 2\mu_{ik}^* \left(Y_{ik} - \mu_{ik}^*\right) + {\mu_{ik}^*}^2\right\} |\tau_i\right] \\
&= \mathrm{E}\left[\left\{\left(Y_{ij} - \mu_{ij}^*\right)^2 + 2\mu_{ij}^* \left(Y_{ij} - \mu_{ij}^*\right) + {\mu_{ij}^*}^2\right\} \left\{\rho^{2(k-j)} \left[\left(Y_{ij} - \mu_{ij}^*\right)^2 - \mu_{ij}^*\right]\right.\right. \\
&\left.\left. + \rho^{k-j}\left(1 - \rho^{k-j}\right)\left(Y_{ij} - \mu_{ij}^*\right) + \mu_{ik}^* + 2\rho^{k-j}\mu_{ik}^*\left(Y_{ij} - \mu_{ij}^*\right) + {\mu_{ik}^*}^2\right\} |\tau_i\right],
\end{aligned}
$$

that can be simplified to

$$
\begin{aligned}
\mathrm{E}(Y_{ij}^2 Y_{ik}^2|\tau_i) &= \mathrm{E}\left[\rho^{2(k-j)}\, Y_{ij}^4 - 2\,\mu_{ij}^*\,\rho^{2(k-j)}\, Y_{ij}^3 - \rho^{2(k-j)}\, Y_{ij}^3 + 2\,\mu_{ik}^*\,\rho^{k-j}\, Y_{ij}^3 + \rho^{k-j}\, Y_{ij}^3\right. \\
&\quad \left. + {\mu_{ij}^*}^2\,\rho^{2(k-j)}\, Y_{ij}^2 - 2\,\mu_{ij}^*\,\mu_{ik}^*\,\rho^{k-j}\, Y_{ij}^2 - \mu_{ij}^*\,\rho^{k-j}\, Y_{ij}^2 + {\mu_{ik}^*}^2\, Y_{ij}^2 + \mu_{ik}^*\, Y_{ij}^2\, |\tau_i\right] \\
&= 2\,{\mu_{ij}^*}^2\,\rho^{2(k-j)} + 4\,{\mu_{ij}^*}^2\,\mu_{ik}^*\,\rho^{k-j} + 2\,\mu_{ij}^*\,\mu_{ik}^*\,\rho^{k-j} + 2\,{\mu_{ij}^*}^2\,\rho^{k-j} \\
&\quad + \mu_{ij}^*\,\rho^{k-j} + {\mu_{ij}^*}^2\,{\mu_{ik}^*}^2 + \mu_{ij}^*\,{\mu_{ik}^*}^2 + {\mu_{ij}^*}^2\,\mu_{ik}^* + \mu_{ij}^*\,\mu_{ik}^*.
\end{aligned}
$$

□

Finally, by substituting $\hat{\psi}(\cdot)$ for $\psi(\cdot)$ in (3.31), we obtain the off-diagonal elements of $\bar{\mathbf{\Omega}}_i$ as:

$$
\begin{aligned}
\mathrm{Cov}\left(Y_{ij}^2, Y_{ik}^2\right) &= 2\rho^{2(k-j)}\bar{\mu}_{ij}^2\, \exp(\sigma_\tau^2) + 4\rho^{k-j}\,\bar{\mu}_{ik}\bar{\mu}_{ij}^2\, \exp(3\sigma_\tau^2) + 2\rho^{k-j}\,\bar{\mu}_{ij}^2\, \exp(\sigma_\tau^2) \\
&\quad + 2\rho^{k-j}\,\bar{\mu}_{ik}\bar{\mu}_{ij}\, \exp(\sigma_\tau^2) + \rho^{k-j}\,\bar{\mu}_{ij} + \bar{\mu}_{ik}^2\bar{\mu}_{ij}^2\, \exp(6\sigma_\tau^2) + \bar{\mu}_{ik}\bar{\mu}_{ij}^2\, \exp(3\sigma_\tau^2) \\
&\quad + \bar{\mu}_{ik}^2\bar{\mu}_{ij}\, \exp(3\sigma_\tau^2) + \bar{\mu}_{ik}\bar{\mu}_{ij}\, \exp(\sigma_\tau^2) - \bar{\lambda}_{ijj}\bar{\lambda}_{ikk} \qquad (3.32)
\end{aligned}
$$

for $j < k;\ j, k = 1, \cdots, n_i$.

3.2.3.1.2 SGQL estimating equation

Now the minimization of Q in (3.24) with regard to σ_τ^2 provides the SGQL estimating equation for σ_τ^2 as

$$\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))}{\partial \sigma_\tau^2} \bar{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\cdot)) \left(\boldsymbol{u}_i - \bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))\right) \quad = \quad 0 \tag{3.33}$$

(Sutradhar, 2004), where $\bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))$ is a vector given by (3.25) and $\bar{\boldsymbol{\Omega}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\cdot))$ is given by (3.27) and (3.32). Also in (3.33),

$$\begin{aligned} \frac{\partial \bar{\boldsymbol{\lambda}}_i^\top}{\partial \sigma_\tau^2} &= \frac{\partial(\bar{\lambda}_{i11}, \ldots, \bar{\lambda}_{ijj}, \ldots, \bar{\lambda}_{in_in_i})}{\partial \sigma_\tau^2}, \text{ with} \\ \frac{\partial \bar{\lambda}_{ijj}}{\partial \sigma_\tau^2} &= \frac{\partial \left(\bar{\mu}_{ij} + \bar{\mu}_{ij}^2 \exp(\sigma_\tau^2)\right)}{\partial \sigma_\tau^2} \\ &= \frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} + 2\bar{\mu}_{ij}\left(\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2}\right) \exp(\sigma_\tau^2) + \bar{\mu}_{ij}^2 \exp(\sigma_\tau^2). \end{aligned} \tag{3.34}$$

Next because $\bar{\mu}_{ij}$ is obtained by replacing $\hat{\psi}(\cdot)$ for $\psi(\cdot)$, it follows from (3.6) and (3.18) that

$$\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} = \bar{\mu}_{ij}\left[\frac{1}{2} + \frac{\partial \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2)}{\partial \sigma_\tau^2}\right]. \tag{3.35}$$

For convenience, by labeling $\bar{\mu}_{ij}$ with $\bar{\mu}_{ij}(z_{ij})$, we then write

$$\begin{aligned} \frac{\partial \hat{\psi}(z_0; \boldsymbol{\beta}, \sigma_\tau^2)}{\partial \sigma_\tau^2} &= -\frac{1}{2} \\ - \quad \exp(\sigma_\tau^2) & \left[\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{\frac{y_{ij} - \bar{\mu}_{ij}(z_0)}{[1+\bar{\mu}_{ij}(z_0)(\exp(\sigma_\tau^2)-1)]^2}\right\}\bar{\mu}_{ij}(z_0)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{\frac{1+y_{ij}(\exp(\sigma_\tau^2)-1)}{[1+\bar{\mu}_{ij}(z_0)(\exp(\sigma_\tau^2)-1)]^2}\right\}\bar{\mu}_{ij}(z_0)}\right], \end{aligned} \tag{3.36}$$

yielding

$$\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} \equiv \frac{\partial \bar{\mu}_{ij}(z_{ij})}{\partial \sigma_\tau^2}$$

$$= \quad -\exp(\sigma_\tau^2) \left[\frac{\sum_{l=1}^{K} \sum_{u=1}^{n_l} w_{lu}(z_{ij}) \left\{ \frac{y_{lu} - \bar{\mu}_{lu}(z_{ij})}{[1+\bar{\mu}_{lu}(z_{ij})(\exp(\sigma_\tau^2)-1)]^2} \right\} \bar{\mu}_{lu}(z_{ij})}{\sum_{l=1}^{K} \sum_{u=1}^{n_l} w_{lu}(z_{ij}) \left\{ \frac{1 + y_{lu}(\exp(\sigma_\tau^2)-1)}{[1+\bar{\mu}_{lu}(z_{ij})(\exp(\sigma_\tau^2)-1)]^2} \right\} \bar{\mu}_{lu}(z_{ij})} \right] \bar{\mu}_{ij}(z_{ij}) \,. \quad (3.37)$$

Notice that in the semi-parametric setup, it is important to accommodate the gradient formula in (3.37), because when $\psi(\cdot)$ is known, $\partial \psi(\cdot)/\partial \sigma_\tau^2 = 0$ and $\partial \mu_{ij}/\partial \sigma_\tau^2 = \frac{1}{2}\mu_{ij}$. Thus, using this later result will produce an inconsistent estimate.

3.2.3.2 SGQL estimation using squared corrected responses

For technical convenience an alternative way to construct a GQL estimating equation for σ_τ^2 would be exploiting the vectors of second order squared corrected responses from the individuals. For the ith individual, let

$$\boldsymbol{g}_i = [(y_{i1} - \bar{\mu}_{i1}(\cdot))^2, \ldots, (y_{ij} - \bar{\mu}_{ij}(\cdot))^2, \ldots, (y_{in_i} - \bar{\mu}_{in_i}(\cdot))^2]^\top$$

denote the second order corrected squared response vector, with known $\bar{\mu}_{ij}(\cdot)$ (3.18) computed from the previous iteration under a suitable iterative scheme. Following (3.33), in this case, we write the SGQL estimating equation for σ_τ^2 as

$$\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\sigma}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))}{\partial \sigma_\tau^2} \bar{\boldsymbol{\Omega}}_{iC}^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\cdot)) \left(\boldsymbol{g}_i - \bar{\boldsymbol{\sigma}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot)) \right) \quad = \quad 0, \qquad (3.38)$$

where

$$\begin{aligned} \bar{\boldsymbol{\sigma}}_i &= \mathrm{E}(\boldsymbol{G}_i) = (\bar{\sigma}_{i11}, \ldots, \bar{\sigma}_{ijj}, \ldots, \bar{\sigma}_{in_in_i})^\top \\ \bar{\boldsymbol{\Omega}}_{iC} &= \mathrm{Cov}(\boldsymbol{G}_i), \end{aligned} \qquad (3.39)$$

with 'C' indicating a 'corrected' response based quantity, and by (3.7)

$$\bar{\sigma}_{ijj} = \bar{\mu}_{ij} + \bar{\mu}_{ij}^2(\exp(\sigma_\tau^2) - 1). \tag{3.40}$$

In (3.38),

$$\begin{aligned}
\frac{\partial \bar{\boldsymbol{\sigma}}_i^\top}{\partial \sigma_\tau^2} &= \frac{\partial}{\partial \sigma_\tau^2}(\bar{\sigma}_{i11}, \ldots, \bar{\sigma}_{ijj}, \ldots, \bar{\sigma}_{in_in_i}), \text{ with} \\
\frac{\partial \bar{\sigma}_{ijj}}{\partial \sigma_\tau^2} &= \frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2} + 2\bar{\mu}_{ij}\left(\frac{\partial \bar{\mu}_{ij}}{\partial \sigma_\tau^2}\right)\big(\exp(\sigma_\tau^2) - 1\big) + \bar{\mu}_{ij}^2 \exp(\sigma_\tau^2), \qquad (3.41)
\end{aligned}$$

where $\partial\bar{\mu}_{ij}/\partial\sigma_\tau^2$ is given by (3.37). Next, as the following lemmas indicate, the formulas for the elements of $\bar{\boldsymbol{\Omega}}_{iC}$ may be computed in a manner similar to that for the elements of $\bar{\boldsymbol{\Omega}}_i$ in (3.33).

Lemma 3.5. *The diagonal elements of* $\bar{\boldsymbol{\Omega}}_{iC}$ *are given by*

$$\begin{aligned}
&\text{Var}\left[(Y_{ij} - \bar{\mu}_{ij})^2\right] = \bar{\mu}_{ij}^4\big(\exp(6\sigma_\tau^2) - 4\exp(3\sigma_\tau^2) + 6\exp(\sigma_\tau^2) - 3\big) \\
&+ \bar{\mu}_{ij}^3\big(6\exp(3\sigma_\tau^2) - 12\exp(\sigma_\tau^2) + 6\big) + \bar{\mu}_{ij}^2\big(7\exp(\sigma_\tau^2) - 4\big) + \bar{\mu}_{ij} - \bar{\sigma}_{ijj}^2, \qquad (3.42)
\end{aligned}$$

and the off-diagonal elements are given by

$$\begin{aligned}
&\text{Cov}\left[(Y_{ij} - \bar{\mu}_{ij})^2, (Y_{ik} - \bar{\mu}_{ik})^2\right] = \left[\bar{\mu}_{ij}^2\bar{\mu}_{ik}\big(4\rho^{k-j} + 1\big) + \bar{\mu}_{ij}\bar{\mu}_{ik}^2\right]\big(\exp(3\sigma_\tau^2) - 2\exp(\sigma_\tau^2) + 1\big) \\
&+ 2\rho^{k-j}\bar{\mu}_{ij}^2\big(\exp(\sigma_\tau^2) - 1 + \rho^{k-j}\exp(\sigma_\tau^2)\big) + \bar{\mu}_{ij}\bar{\mu}_{ik}\left[2\rho^{k-j}\big(\exp(\sigma_\tau^2) - 1\big) + \exp(\sigma_\tau^2)\right] \\
&+ \rho^{k-j}\bar{\mu}_{ij} + \bar{\mu}_{ij}^2\bar{\mu}_{ik}^2\big(\exp(6\sigma_\tau^2) - 4\exp(3\sigma_\tau^2) + 6\exp(\sigma_\tau^2) - 3\big) - \bar{\sigma}_{ijj}\bar{\sigma}_{ikk}. \qquad (3.43)
\end{aligned}$$

Proof. The equation (3.42) can be easily derived by noting that conditional on τ_i, Y_{ij} follows a Poisson distribution [Sutradhar (2011, Section 6.3.1)]. Now to obtain the result in (3.43), for $\forall\ 1 \leq j < k \leq n_i$, by Lemma (3.3), we first obtain the conditional expectation as

$$\begin{aligned}
&\mathrm{E}\left[\left(Y_{ik}-\mu_{ik}\right)^2\left(Y_{ij}-\mu_{ij}\right)^2\,\middle|\,\tau_i\right]=\mathrm{E}\left[\left(Y_{ik}-\mu^*_{ik}+\mu^*_{ik}-\mu_{ik}\right)^2\left(Y_{ij}-\mu^*_{ij}+\mu^*_{ij}-\mu_{ij}\right)^2\,\middle|\,\tau_i\right]\\
&=\ \mathrm{E}\left[\left\{\left(Y_{ik}-\mu^*_{ik}\right)^2+2\left(\mu^*_{ik}-\mu_{ik}\right)\left(Y_{ik}-\mu^*_{ik}\right)+\left(\mu^*_{ik}-\mu_{ik}\right)^2\right\}\cdot\right.\\
&\qquad\left.\left\{\left(Y_{ij}-\mu^*_{ij}\right)^2+2\left(\mu^*_{ij}-\mu_{ij}\right)\left(Y_{ij}-\mu^*_{ij}\right)+\left(\mu^*_{ij}-\mu_{ij}\right)^2\right\}\,\middle|\,\tau_i\right]\\
&=\ \mathrm{E}\left[\left\{\rho^{2(k-j)}\left[\left(Y_{ij}-\mu^*_{ij}\right)^2-\mu^*_{ij}\right]+\rho^{k-j}\left(1-\rho^{k-j}\right)\left(Y_{ij}-\mu^*_{ij}\right)+\mu^*_{ik}+\left(\mu^*_{ik}-\mu_{ik}\right)^2\right.\right.\\
&\qquad\left.\left.+\,2\rho^{k-j}\left(\mu^*_{ik}-\mu_{ik}\right)\left(Y_{ij}-\mu^*_{ij}\right)\right\}\left\{\left(Y_{ij}-\mu^*_{ij}\right)^2+2\left(\mu^*_{ij}-\mu_{ij}\right)\left(Y_{ij}-\mu^*_{ij}\right)+\left(\mu^*_{ij}-\mu_{ij}\right)^2\right\}\,\middle|\,\tau_i\right]\\
&=\ \mathrm{E}\left[\left\{\rho^{2(k-j)}\left(Y_{ij}-\mu^*_{ij}\right)^2+\rho^{k-j}\left[1-\rho^{k-j}+2\left(\mu^*_{ik}-\mu_{ik}\right)\right]\left(Y_{ij}-\mu^*_{ij}\right)\right.\right.\\
&\qquad\left.+\left[\mu^*_{ik}+\left(\mu^*_{ik}-\mu_{ik}\right)^2-\rho^{2(k-j)}\mu^*_{ij}\right]\right\}\\
&\qquad\left.\left\{\left(Y_{ij}-\mu^*_{ij}\right)^2+2\left(\mu^*_{ij}-\mu_{ij}\right)\left(Y_{ij}-\mu^*_{ij}\right)+\left(\mu^*_{ij}-\mu_{ij}\right)^2\right\}\,\middle|\,\tau_i\right]\\
&=\ \mathrm{E}\left[\rho^{2(k-j)}\left(Y_{ij}-\mu^*_{ij}\right)^4+\left\{2\left(\mu^*_{ij}-\mu_{ij}\right)\rho^{2(k-j)}+\rho^{k-j}\left[1-\rho^{k-j}+2\left(\mu^*_{ik}-\mu_{ik}\right)\right]\right\}\left(Y_{ij}-\mu^*_{ij}\right)^3\right.\\
&\qquad+\left(Y_{ij}-\mu^*_{ij}\right)^2\left\{\rho^{2(k-j)}\left(\mu^*_{ij}-\mu_{ij}\right)^2+\mu^*_{ik}+\left(\mu^*_{ik}-\mu_{ik}\right)^2-\rho^{2(k-j)}\mu^*_{ij}\right.\\
&\qquad\left.+2\left(\mu^*_{ij}-\mu_{ij}\right)\rho^{k-j}\left[1-\rho^{k-j}+2\left(\mu^*_{ik}-\mu_{ik}\right)\right]\right\}\\
&\qquad+\left(Y_{ij}-\mu^*_{ij}\right)\left\{2\left(\mu^*_{ij}-\mu_{ij}\right)\left[\mu^*_{ik}+\left(\mu^*_{ik}-\mu_{ik}\right)^2-\rho^{2(k-j)}\mu^*_{ij}\right]\right.\\
&\qquad\left.+\rho^{k-j}\left[1-\rho^{k-j}+2\left(\mu^*_{ik}-\mu_{ik}\right)\right]\left(\mu^*_{ij}-\mu_{ij}\right)^2\right\}\\
&\qquad\left.+\left(\mu^*_{ij}-\mu_{ij}\right)^2\left[\mu^*_{ik}+\left(\mu^*_{ik}-\mu_{ik}\right)^2-\rho^{2(k-j)}\mu^*_{ij}\right]\,\middle|\,\tau_i\right]\\
&=\ 4\rho^{k-j}\mu^*_{ik}{\mu^*_{ij}}^2-4\rho^{k-j}\mu_{ik}{\mu^*_{ij}}^2+2\rho^{k-j}{\mu^*_{ij}}^2-4\rho^{k-j}\mu_{ij}\mu^*_{ik}\mu^*_{ij}\\
&\ +2\rho^{k-j}\mu^*_{ik}\mu^*_{ij}+4\rho^{k-j}\mu_{ik}\mu_{ij}\mu^*_{ij}-2\rho^{k-j}\mu_{ij}\mu^*_{ij}\\
&\ -2\rho^{k-j}\mu_{ik}\mu^*_{ij}+\rho^{k-j}\mu^*_{ij}+2\rho^{2(k-j)}{\mu^*_{ij}}^2+{\mu^*_{ik}}^2{\mu^*_{ij}}^2-2\mu_{ik}\mu^*_{ik}{\mu^*_{ij}}^2\\
&\ +\mu^*_{ik}{\mu^*_{ij}}^2+\mu^2_{ik}{\mu^*_{ij}}^2-2\mu_{ij}{\mu^*_{ik}}^2\mu^*_{ij}+{\mu^*_{ik}}^2\mu^*_{ij}\\
&\ +4\mu_{ik}\mu_{ij}\mu^*_{ik}\mu^*_{ij}-2\mu_{ij}\mu^*_{ik}\mu^*_{ij}-2\mu_{ik}\mu^*_{ik}\mu^*_{ij}+\mu^*_{ik}\mu^*_{ij}\\
&\ -2\mu^2_{ik}\mu_{ij}\mu^*_{ij}+\mu^2_{ik}\mu^*_{ij}+\mu^2_{ij}{\mu^*_{ik}}^2-2\mu_{ik}\mu^2_{ij}\mu^*_{ik}+\mu^2_{ij}\mu^*_{ik}+\mu^2_{ik}\mu^2_{ij}.
\end{aligned}$$

Then, by averaging over the distribution of $\tau_i \overset{iid}{\sim} N(0,1)$, we obtain the unconditional expectation as

$$
\begin{aligned}
&\mathrm{E}\left[(Y_{ik}-\mu_{ik})^2(Y_{ij}-\mu_{ij})^2\right] = 4\rho^{k-j}\mu_{ik}{\mu_{ij}}^2 e^{3\sigma_\tau^2} - 4\rho^{k-j}\mu_{ik}{\mu_{ij}}^2 e^{\sigma_\tau^2} + 2\rho^{k-j}{\mu_{ij}}^2 e^{\sigma_\tau^2} - 4\rho^{k-j}\mu_{ik}\mu_{ij}^2 e^{\sigma_\tau^2} \\
&+ 2\rho^{k-j}\mu_{ik}\mu_{ij}e^{\sigma_\tau^2} + 4\rho^{k-j}\mu_{ik}\mu_{ij}^2 - 2\rho^{k-j}\mu_{ij}^2 - 2\rho^{k-j}\mu_{ik}\mu_{ij} + \rho^{k-j}\mu_{ij} \\
&+ 2\rho^{2(k-j)}\mu_{ij}^2 e^{\sigma_\tau^2} + \mu_{ik}^2\mu_{ij}^2 e^{6\sigma_\tau^2} - 2\mu_{ik}^2\mu_{ij}^2 e^{3\sigma_\tau^2} \\
&+ \mu_{ik}\mu_{ij}^2 e^{3\sigma_\tau^2} + \mu_{ik}^2\mu_{ij}^2 e^{\sigma_\tau^2} - 2\mu_{ik}^2\mu_{ij}^2 e^{3\sigma_\tau^2} + \mu_{ik}^2\mu_{ij}e^{3\sigma_\tau^2} \\
&+ 4\mu_{ik}^2\mu_{ij}^2 e^{\sigma_\tau^2} - 2\mu_{ik}\mu_{ij}^2 e^{\sigma_\tau^2} - 2\mu_{ik}^2\mu_{ij}e^{\sigma_\tau^2} + \mu_{ik}\mu_{ij}e^{\sigma_\tau^2} \\
&- 2\mu_{ik}^2\mu_{ij}^2 + \mu_{ik}^2\mu_{ij} + \mu_{ik}^2\mu_{ij}^2 e^{\sigma_\tau^2} - 2\mu_{ik}^2\mu_{ij}^2 + \mu_{ik}\mu_{ij}^2 + \mu_{ik}^2\mu_{ij}^2 \\
&= \mu_{ik}\mu_{ij}^2\left[4\rho^{k-j}\left(e^{3\sigma_\tau^2} - 2e^{\sigma_\tau^2} + 1\right) + e^{3\sigma_\tau^2} - 2e^{\sigma_\tau^2} + 1\right] + 2\rho^{k-j}\mu_{ij}^2\left(e^{\sigma_\tau^2} - 1 + \rho^{k-j}e^{\sigma_\tau^2}\right) \\
&+ \mu_{ik}\mu_{ij}\left[2\rho^{k-j}\left(e^{\sigma_\tau^2} - 1\right) + e^{\sigma_\tau^2}\right] + \rho^{k-j}\mu_{ij} \\
&+ \mu_{ik}^2\mu_{ij}^2\left(e^{6\sigma_\tau^2} - 4e^{3\sigma_\tau^2} + 6e^{\sigma_\tau^2} - 3\right) + \mu_{ik}^2\mu_{ij}\left(e^{3\sigma_\tau^2} - 2e^{\sigma_\tau^2} + 1\right) \\
&= \mu_{ik}\mu_{ij}^2\left(4\rho^{k-j} + 1\right)\left(e^{3\sigma_\tau^2} - 2e^{\sigma_\tau^2} + 1\right) + 2\rho^{k-j}\mu_{ij}^2\left(e^{\sigma_\tau^2} - 1 + \rho^{k-j}e^{\sigma_\tau^2}\right) \\
&+ \mu_{ik}\mu_{ij}\left[2\rho^{k-j}\left(e^{\sigma_\tau^2} - 1\right) + e^{\sigma_\tau^2}\right] \\
&+ \rho^{k-j}\mu_{ij} + \mu_{ik}^2\mu_{ij}^2\left(e^{6\sigma_\tau^2} - 4e^{3\sigma_\tau^2} + 6e^{\sigma_\tau^2} - 3\right) + \mu_{ik}^2\mu_{ij}\left(e^{3\sigma_\tau^2} - 2e^{\sigma_\tau^2} + 1\right) \\
&= \left[\mu_{ik}\mu_{ij}^2\left(4\rho^{k-j} + 1\right) + \mu_{ik}^2\mu_{ij}\right]\left(e^{3\sigma_\tau^2} - 2e^{\sigma_\tau^2} + 1\right) + 2\rho^{k-j}\mu_{ij}^2\left(e^{\sigma_\tau^2} - 1 + \rho^{k-j}e^{\sigma_\tau^2}\right) \\
&\quad + \mu_{ik}\mu_{ij}\left[2\rho^{k-j}\left(e^{\sigma_\tau^2} - 1\right) + e^{\sigma_\tau^2}\right] + \rho^{k-j}\mu_{ij} + \mu_{ik}^2\mu_{ij}^2\left(e^{6\sigma_\tau^2} - 4e^{3\sigma_\tau^2} + 6e^{\sigma_\tau^2} - 3\right).
\end{aligned}
$$

Finally, by combining terms and substituting $\hat{\psi}(\cdot)$ for $\psi(\cdot)$, we obtain (3.43). □

3.2.3.3 Normal approximation based SGQL estimation using squared corrected responses

In Sections 3.2.3.1 and 3.2.3.2 we computed the covariance matrix for the raw and squared corrected responses for the construction of the estimating equation for σ_τ^2.

Notice that these computations were done by retaining the original (count) distribution nature of the responses. However, there exists an alternative approach (Zhao and Prentice, 1990, Prentice and Zhao, 1991) where this type of fourth moments calculations are done by changing the distributional assumption for the responses. More specifically, these authors proceed as if the repeated count responses are multivariate normal random variables in order to compute these higher order moments, while using the correct means and variances.

This "normality" based SGQL estimating equation would be the same as that of (3.38) constructed based on squared corrected responses except that the fourth order moment matrix $\bar{\boldsymbol{\Omega}}_{iC}$ is now replaced with a normality based fourth order moment matrix, say $\bar{\boldsymbol{\Omega}}_{iC,N}$. Thus, in notation of (3.38),

$$\mathrm{Cov}_N(\boldsymbol{G}_i) = \bar{\boldsymbol{\Omega}}_{iC,N}, \tag{3.44}$$

where the elements of this matrix are computed from the normality based fourth order product moments formula

$$\begin{aligned} & \mathrm{E}_N\left[(Y_{ij}-\bar{\mu}_{ij})(Y_{ik}-\bar{\mu}_{ik})(Y_{il}-\bar{\mu}_{il})(Y_{im}-\bar{\mu}_{im})\right] \\ = \; & \bar{\sigma}_{ijk}\bar{\sigma}_{ilm}+\bar{\sigma}_{ijl}\bar{\sigma}_{ikm}+\bar{\sigma}_{ijm}\bar{\sigma}_{ikl}. \end{aligned} \tag{3.45}$$

for $i = 1, \ldots, K$ and $1 \leq j, k, l, m \leq n_i$. For example, under normality,

$$\begin{aligned} \mathrm{Var}\left[(Y_{ij}-\bar{\mu}_{ij})^2\right] &= E_N\left[(Y_{ij}-\bar{\mu}_{ij})^4\right] - \bar{\sigma}_{ijj}^2 \\ &= 3\bar{\sigma}_{ijj}^2 - \bar{\sigma}_{ijj}^2 = 2\bar{\sigma}_{ijj}^2, \end{aligned} \tag{3.46}$$

by (3.45). Notice that the normality assumption for count responses $\{y_{ij}, j = 1, \ldots, n_i\}$

simplifies the computation of high order moments. Also we remark that this approximation appears to work well for repeated count data in the GLM setup (Sutradhar, 2011, Chapter 8). In this section, we have considered its use in the semi-parametric longitudinal mixed model setup. The finite sample performance of this approach in the present setup will be given in Section 3.4.

For completeness, using the notations from (3.44)−(3.46), we now write the desired normality based SGQL estimating equation for σ_τ^2 as

$$\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\sigma}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))}{\partial \sigma_\tau^2} \bar{\boldsymbol{\Omega}}_{iC,N}^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\cdot)) \left(\boldsymbol{g}_i - \bar{\boldsymbol{\sigma}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))\right) = 0, \quad (3.47)$$

which is solved iteratively until convergence.

3.2.4 Estimation of the longitudinal correlation index parameter ρ

The estimation of the regression parameter $\boldsymbol{\beta}$ and over-dispersion parameter σ_τ^2 are discussed in Sections 3.2.2 and 3.2.3, respectively. Notice that their estimation requires the longitudinal correlation index parameter ρ to be known. We show in this section that the ρ parameter can be estimated by solving an unbiased moment equation that leads to a consistent estimator. For the purpose, it follows from (3.7) and (3.8) that the variances and the lag 1 covariances of the repeated counts under the present model have the formulas

$$\begin{aligned} \mathrm{E}\left[(Y_{ij} - \mu_{ij})^2\right] &= \sigma_{ijj} = \mu_{ij} + \mu_{ij}^2 \left(\exp(\sigma_\tau^2) - 1\right), \text{ and} \\ \mathrm{E}\left[(Y_{ij} - \mu_{ij})(Y_{i,j+1} - \mu_{i,j+1})\right] &= \sigma_{i,j,j+1} = \rho\mu_{ij} + \mu_{ij}\mu_{i,j+1} \left(\exp(\sigma_\tau^2) - 1\right), \quad (3.48) \end{aligned}$$

respectively. Let $y_{ij}^* = (y_{ij} - \mu_{ij}) / (\sigma_{ijj})^{1/2}$. It is then straightforward to observe that

$$\mathrm{E}\left[\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} {y_{ij}^*}^2}{\sum_{i=1}^{K} n_i}\right] = 1, \tag{3.49}$$

and

$$\begin{aligned}\mathrm{E}\left[\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1} y_{ij}^* y_{i,j+1}^*}{\sum_{i=1}^{K}(n_i-1)}\right] &= \frac{\rho\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\frac{\mu_{ij}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^{K}(n_i-1)} \\ &+ \frac{(\exp(\sigma_\tau^2)-1)\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\frac{\mu_{ij}\mu_{i,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^{K}(n_i-1)}. \end{aligned} \tag{3.50}$$

Now by exploiting (3.49) and (3.50), more specifically considering the ratio of the quantities within the square brackets in (3.50) and (3.49) and denoting it by a_1, that is,

$$a_1 = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1} y_{ij}^* y_{i,j+1}^* / \sum_{i=1}^{K}(n_i-1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} {y_{ij}^*}^2 / \sum_{i=1}^{K} n_i}, \tag{3.51}$$

we may then write a first order approximate expectation as

$$\mathrm{E}[a_1] \approx \rho g_1 + b_1, \tag{3.52}$$

where

$$b_1 = \left(\exp\left(\sigma_\tau^2\right) - 1\right)\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\phi_{ij}\phi_{i,j+1} \Big/ \sum_{i=1}^{K}(n_i-1), \tag{3.53}$$

and

$$g_1 = \sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\mu_{ij}\left(\sigma_{ijj}\sigma_{i,j+1,j+1}\right)^{-1/2} \Big/ \sum_{i=1}^{K}(n_i-1), \tag{3.54}$$

with $\phi_{ij} = \mu_{ij} / (\sigma_{ijj})^{1/2}$.

Next by replacing $\mu_{ij},\ \sigma_{ijj}$, and $\sigma_{i,j+1,j+1}$ in (3.51), (3.53) and (3.54) with $\bar{\mu}_{ij},\ \bar{\sigma}_{ijj}$, and $\bar{\sigma}_{i,j+1,j+1}$ respectively, one can obtain $\bar{a}_1, \bar{b}_1,$ and $\bar{g}_1$, from $a_1, b_1,$ and g_1, respectively. Consequently, from (3.52), we write an approximate moment estimator of ρ as

$$\hat{\rho} \ = \ \frac{\bar{a}_1 - \bar{b}_1}{\bar{g}_1}. \tag{3.55}$$

Note that the overall estimation for all functions and parameters, that is the estimation of the nonparametric function $\psi(\cdot)$ (Section 3.2.1), regression effects $\boldsymbol{\beta}$ (Section 3.2.2), over-dispersion component σ^2_τ (Section 3.2.3), and the longitudinal correlation index parameter ρ (Section 3.2.4), is carried out in iterated stages until convergence.

3.3 Asymptotic results

For the definition of the notations such as o, O, o_p and O_p used in this thesis, we refer to Bishop et al. (2007), Chapter 14.

3.3.1 Consistency of the SQL estimator of $\psi(\cdot)$

Note that the SQL estimating equation (3.15) is an extension of the well known QL estimating equation (Wedderburn, 1974). This estimating equation, which is free of ρ, is written by exploiting the means and the variances of the responses, variance being a function of the mean in the present GLMM setup, by treating the repeated responses of an individual as independent. $\psi(z_{\ell u})$ has to be evaluated for all $u = 1, \ldots, n_l$; and $\ell = 1, \ldots, K$. For convenience, in (3.15), we have shown the estimation for $\psi(z_0)$ for $z_0 =$

$z_{\ell u}$ for a selected value of ℓ and u. This estimate for $\psi(z_0)$ was denoted by $\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2)$. For notational simplicity, here we use $\mu_{ij}(z_0)$ for $\mu_{ij}(\boldsymbol{\beta},\sigma_\tau^2,\psi(z_0))$. Now, for known $\boldsymbol{\beta}$ and σ_τ^2, and for true mean $\mu_{ij} = \exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \frac{\sigma_\tau^2}{2} + \psi(z_{ij}))$, a Taylor expansion of (3.15) around $\psi(z_0)$ gives

$$\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2) - \psi(z_0) = A_K + H_K + O(|\hat{\psi}(z_0;\boldsymbol{\beta},\sigma_\tau^2) - \psi(z_0)|^2) \tag{3.56}$$

where

$$A_K = \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\,\frac{y_{ij}-\mu_{ij}}{1+\mu_{ij}(z_0)\,(\exp(\sigma_\tau^2)-1)}, \quad \text{and}$$

$$H_K = \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\,\frac{\mu_{ij}-\mu_{ij}(z_0)}{1+\mu_{ij}(z_0)\,(\exp(\sigma_\tau^2)-1)}$$

with $B_K = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\,\frac{\mu_{ij}(z_0)}{1+\mu_{ij}(z_0)(\exp(\sigma_\tau^2)-1)}$, and $p_{ij}(z_0)$ is the short abbreviation for $p_{ij}(\frac{z_0-z_{ij}}{b})$ defined in (3.12), b being the so-called bandwidth parameter. Because it is easy to show that A_K has zero mean and bounded variance, one may then write

$$A_K = O_p(1/\sqrt{K}) \tag{3.57}$$

as $K \to \infty$, according to Theorem 14.4-1 in Bishop et al. (2007). Now we show that H_K approaches zero in the order of $O(b^2)$.

Lemma 3.6. *The kernel density $p_{ij}(z_0)$ defined by (3.13)–(3.14) has the expectation given by*

$$\mathrm{E}\left[p_{ij}(z_0)\,(z_{ij}-z_0)|\mathbf{x}_{ij}\right] \;=\; O(b^2) \tag{3.58}$$

as $K \to \infty$.

Proof. Let $h(z_{ij}; \mathbf{x}_{ij})$ be the pdf of z_{ij} conditional on $\mathbf{x}_{ij}$, then

$$\begin{aligned}
\mathrm{E}\left[p_{ij}(z_0)\,(z_{ij}-z_0)|\mathbf{x}_{ij}\right] &= \int p_{ij}(z_0)\,(z_{ij}-z_0)\,h(z_{ij};\mathbf{x}_{ij})\,dz_{ij} \\
&= \int p_{ij}(z_0)\left[(z_{ij}-z_0)\,h(z_0;\mathbf{x}_{ij}) + O\left((z_{ij}-z_0)^2\right)\right]\,dz_{ij} \\
&\qquad\qquad \text{since } h(z_{ij};\mathbf{x}_{ij}) = h(z_0;\mathbf{x}_{ij}) + O(z_{ij}-z_0) \\
&= h(z_0;\mathbf{x}_{ij}) \int p_{ij}(z_0)\,(z_{ij}-z_0)\,dz_{ij} + O(b^2), \qquad (3.59)
\end{aligned}$$

because $\int p_{ij}(z_0)\,O\left((z_{ij}-z_0)^2\right)\,dz_{ij}$ can be shown bounded in the order of b^2 ($O(b^2)$). Next the first term in (3.59) gives zero as $p_{ij}(z_0)$ is symmetric about z_0, yielding the lemma. □

Lemma 3.7. *The quantity* H_K *in (3.56) satisfies*

$$H_K \;=\; O(b^2) \qquad (3.60)$$

as $K \to \infty$.

Proof. By a first order Taylor expansion, we may write

$$H_K \simeq \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\,\frac{\mu_{ij}(z_0)\,\psi'(z_0)}{1+\mu_{ij}(z_0)\,(\exp(\sigma_\tau^2)-1)}\,(z_{ij}-z_0)\,.$$

Next, we rewrite this H_K as

$$\begin{aligned}
H_K = &\frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\mu_{ij}(z_0)\,\psi'(z_0)}{1+\mu_{ij}(z_0)\,(\exp(\sigma_\tau^2)-1)}\left\{p_{ij}(z_0)\,(z_{ij}-z_0) - \mathrm{E}\left[p_{ij}(z_0)\,(z_{ij}-z_0)|\mathbf{x}_{ij}\right]\right\} \\
&+ \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\mu_{ij}(z_0)\,\psi'(z_0)}{1+\mu_{ij}(z_0)\,(\exp(\sigma_\tau^2)-1)}\,\mathrm{E}\left[p_{ij}(z_0)\,(z_{ij}-z_0)|\mathbf{x}_{ij}\right].
\end{aligned}$$

Here the second term has the order $O(b^2)$ by Lemma 3.6. For the first term, due to $p_{ij}(z_0)$, its variance is in the order of $O(b^2/K)$, so it has the order $O_p(b/\sqrt{K})$, which can be neglected. Thus H_K has the order $O(b^2)$. $\square$

We now apply Lemma 3.7 and use (3.57) in (3.56). Thus, we write

$$\hat{\psi}(z_0; \boldsymbol{\beta}, \sigma_\tau^2) - \psi(z_0) = A_K + O(b^2) + o_p(1/\sqrt{K}) = O_p(1/\sqrt{K}) + O(b^2) \quad (3.61)$$

as $K \to \infty$, where $b \propto K^{-\alpha}$ (Pagan and Ullah, 1999, Horowitz, 2009) for a suitable value for $\alpha \in [1/5, 1/3]$ (Lin and Carroll, 2001). Notice that when $\alpha > 1/4$, $\hat{\psi}(z_0; \boldsymbol{\beta}, \sigma_\tau^2)$ is a consistent estimator of nonparametric function value $\psi(z_0)$.

3.3.2 Consistency of the SGQL estimator of $\boldsymbol{\beta}$ and its asymptotic multivariate normal distribution

The SGQL estimator of $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}$, is obtained by (3.20). Before we derive the asymptotic properties of the estimator $\hat{\boldsymbol{\beta}}$, it is convenient to prove the following two lemmas. The main result is given in Theorem 3.1.

First notice that in the SGQL estimating equation (3.20) for $\boldsymbol{\beta}$, we have used $\hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)$ for $\hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2)$ by suppressing z_{ij} for notational simplicity. Let

$$\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2) \equiv [\hat{\psi}(z_{i1}; \boldsymbol{\beta}, \sigma_\tau^2), \ldots, \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2), \ldots, \hat{\psi}(z_{in_i}; \boldsymbol{\beta}, \sigma_\tau^2)].$$

Also recall from (3.20) that $\hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)$ was used to define $\bar{\boldsymbol{\mu}}_i(\cdot)$ and $\bar{\boldsymbol{\Sigma}}_i$ from $\boldsymbol{\mu}_i(\cdot)$ and $\boldsymbol{\Sigma}_i$ respectively. Here $\hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)$ refers to using all values of $\hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2)$ for $j = 1, \ldots, n_i$. We now refer to (3.20) and express its solution as in the following lemma.

Lemma 3.8. *Suppose that the estimating equation (3.20) is written as* $KD_K(\boldsymbol{\beta}) = 0$, *where*

$$\begin{aligned}\boldsymbol{D}_K(\boldsymbol{\beta}) &= \frac{1}{K}\sum_{i=1}^{K}\frac{\partial\bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta},\sigma_\tau^2,\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))}{\partial\boldsymbol{\beta}}\bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2),\rho)\\ &\times \left[\boldsymbol{Y}_i-\bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta},\sigma_\tau^2,\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))\right].\end{aligned}$$

Then we can write

$$\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} = [\boldsymbol{F}(\boldsymbol{\beta})]^{-1}\boldsymbol{D}_K(\boldsymbol{\beta})+o_p(1/\sqrt{K}), \tag{3.62}$$

where

$$\boldsymbol{F}(\boldsymbol{\beta})=\mathrm{E}\left[\frac{\partial\bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta},\sigma_\tau^2,\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))}{\partial\boldsymbol{\beta}}\bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2),\rho)\frac{\partial\bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta},\sigma_\tau^2,\hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))}{\partial\boldsymbol{\beta}^\top}\right].$$

Proof. Because the SGQL estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ obtained from (3.20) satisfies $\boldsymbol{D}_K(\hat{\boldsymbol{\beta}})=0$, a linear (first order) Taylor expansion about true $\boldsymbol{\beta}$ provides

$$\boldsymbol{D}_K(\boldsymbol{\beta})+\boldsymbol{D}'_K(\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})+O(|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}|^2)=0, \tag{3.63}$$

yielding

$$\begin{aligned}\hat{\boldsymbol{\beta}}-\boldsymbol{\beta} &= -[\boldsymbol{D}'_K(\boldsymbol{\beta})]^{-1}[\boldsymbol{D}_K(\boldsymbol{\beta})+O(|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}|^2)]\\ &= [\boldsymbol{F}_K(\boldsymbol{\beta})]^{-1}\boldsymbol{D}_K(\boldsymbol{\beta})+o_p(1/\sqrt{K}), \end{aligned}\tag{3.64}$$

where

$$\boldsymbol{F}_K(\boldsymbol{\beta})=-\boldsymbol{D}'_K(\boldsymbol{\beta})=-\frac{\partial\boldsymbol{D}_K(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\top}$$

$$= \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2), \rho) \frac{\partial \bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \boldsymbol{\beta}^\top}.$$

Here we in fact applied Lemma 3.9 and $b \propto K^{-\alpha}$ with $1/5 \leq \alpha \leq 1/3$ (Lin and Carroll, 2001) to learn that $O(|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}|^2)$ is in the order of $o_p(1/\sqrt{K})$.

Now under the assumption that

$$\begin{aligned}\lim_{K\to\infty} \boldsymbol{F}_K(\boldsymbol{\beta}) &= \mathrm{E}\left[\frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2), \rho) \frac{\partial \bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \boldsymbol{\beta}^\top}\right] \\ &= \boldsymbol{F}(\boldsymbol{\beta}),\end{aligned}$$

the lemma follows. □

Lemma 3.9. *$\boldsymbol{D}_K(\boldsymbol{\beta})$ in Lemma 3.8 can be written as*

$$\boldsymbol{D}_K(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K}(\bar{\boldsymbol{Z}}_{1i} - \bar{\boldsymbol{Z}}_{2i})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + O(b^2) + o_p(1/\sqrt{K}), \tag{3.65}$$

where

$$\bar{\boldsymbol{Z}}_{1i} = \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho),$$

and

$$\bar{\boldsymbol{Z}}_{2i} = \left(\bar{\boldsymbol{Z}}_{2i1}, \cdots, \bar{\boldsymbol{Z}}_{2in_i}\right)$$

with

$$\begin{aligned}\bar{\boldsymbol{Z}}_{2ij} &= \sum_{i'=1}^{K}\sum_{j'=1}^{n_i}\sum_{k'=1}^{n_i} \frac{1}{B_K(z_{i'k'})} \frac{\partial \bar{\mu}_{i'j'}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{i'j'}; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{v}_{1i'}^{j'k'}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho) \cdot \\ &\qquad \mu_{i'k'}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{i'k'})) \frac{p_{ij}(z_{i'k'})}{1 + \mu_{ij}(z_{i'k'})\left(e^{\sigma_\tau^2} - 1\right)},\end{aligned}$$

where $\bar{v}_{1i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)$ is the (j,k)th element of $\bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)$.

Proof. Write $\boldsymbol{D}_K(\boldsymbol{\beta})$ as

$$\begin{aligned}
\boldsymbol{D}_K(\boldsymbol{\beta}) &= \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2), \rho) \left[\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right] \\
&\quad - \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2), \rho) \cdot \\
&\qquad \left[\bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2)) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right] \\
&= \boldsymbol{D}_{1K}(\boldsymbol{\beta}) - \boldsymbol{D}_{2K}(\boldsymbol{\beta}),
\end{aligned}$$

where

$$\boldsymbol{D}_{1K}(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K} \bar{\boldsymbol{Z}}_{1i}\left[\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right], \tag{3.66}$$

and

$$\begin{aligned}
\boldsymbol{D}_{2K}(\boldsymbol{\beta}) &= \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2), \rho) \cdot \\
&\qquad \left[\bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i;\boldsymbol{\beta},\sigma_\tau^2)) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right] \\
&= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} \frac{\partial \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{v}_{1i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho) \cdot \\
&\qquad \left[\bar{\mu}_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ik};\boldsymbol{\beta},\sigma_\tau^2)) - \mu_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik}))\right] \\
&= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} \left\{ \frac{\partial \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{v}_{1i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)\mu_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik})) \cdot \right. \\
&\qquad \left. \left[\hat{\psi}(z_{ik};\boldsymbol{\beta},\sigma_\tau^2) - \psi(z_{ik})\right] + O_p\left(\left[\hat{\psi}(z_{ik};\boldsymbol{\beta},\sigma_\tau^2) - \psi(z_{ik})\right]^2\right)\right\} \\
&= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} \frac{\partial \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{v}_{1i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)\mu_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik})) \cdot \\
&\qquad \left[\hat{\psi}(z_{ik};\boldsymbol{\beta},\sigma_\tau^2) - \psi(z_{ik})\right] + o_p(1/\sqrt{K}) \qquad \text{by (3.61)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} \frac{\partial \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{v}_{1i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)\mu_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik}))\cdot \\
&\left[\frac{1}{K}\frac{1}{B_K(z_{ik})}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}} p_{i'j'}(z_{ik})\, \frac{Y_{i'j'}-\mu_{i'j'}}{1+\mu_{i'j'}(z_{ik})\,(e^{\sigma_\tau^2}-1)} + O(b^2) + o_p(1/\sqrt{K})\right] + o_p(1/\sqrt{K}) \\
&= \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}}\left[\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i}\frac{1}{B_K(z_{ik})}\frac{\partial \bar{\mu}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))}{\partial \boldsymbol{\beta}} \bar{v}_{1i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)\mu_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik}))\right. \\
&\left.\frac{p_{i'j'}(z_{ik})}{1+\mu_{i'j'}(z_{ik})\,(e^{\sigma_\tau^2}-1)}\right](Y_{i'j'}-\mu_{i'j'}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{i'j'}))) + O(b^2) + o_p(1/\sqrt{K}) \\
&= \frac{1}{K}\sum_{i=1}^{K}\bar{\boldsymbol{Z}}_{2i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))) + O(b^2) + o_p(1/\sqrt{K}). \qquad (3.67)
\end{aligned}
$$

The above results for $\boldsymbol{D}_{1K}(\boldsymbol{\beta})$ (3.66) and $\boldsymbol{D}_{2K}(\boldsymbol{\beta})$ (3.67) together complete the proof.

□

Let $\boldsymbol{Z}_{1i}$ and $\boldsymbol{Z}_{2i}$ be the $\bar{\boldsymbol{Z}}_{1i}$ and $\bar{\boldsymbol{Z}}_{2i}$ defined in Lemma 3.9, respectively, with $\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)$ being replaced by its true value $\psi(z_{ij})$ for all j, and $\partial\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)/\partial\boldsymbol{\beta}$ given in (3.24) being replaced by

$$
\begin{aligned}
&\frac{\partial \tilde{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)}{\partial \boldsymbol{\beta}} \qquad (3.68) \\
&= \frac{-\frac{1}{K}\sum_{i=1}^{K}\sum_{l=1}^{n_i} w_{il}(z_{ij})\mathrm{E}\left[\left\{\frac{1+y_{il}\left(\exp(\sigma_\tau^2)-1\right)}{\left[1+\bar{\mu}_{il}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))(\exp(\sigma_\tau^2)-1)\right]^2}\right\}\bar{\mu}_{il}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))\right]\boldsymbol{x}_{il}}{\frac{1}{K}\sum_{i=1}^{K}\sum_{l=1}^{n_i} w_{il}(z_{ij})\mathrm{E}\left[\left\{\frac{1+y_{il}(\exp(\sigma_\tau^2)-1)}{\left[1+\bar{\mu}_{il}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))(\exp(\sigma_\tau^2)-1)\right]^2}\right\}\bar{\mu}_{il}(\boldsymbol{\beta},\sigma_\tau^2,\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2))\right]}
\end{aligned}
$$

for all j. Then a linear Taylor expansion of $\bar{\boldsymbol{Z}}_{1i}-\bar{\boldsymbol{Z}}_{2i}$ with respect to $\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)$ at point $\psi(z_{ij})$, and $\partial\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)/\partial\boldsymbol{\beta}$ at point $\partial\tilde{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)/\partial\boldsymbol{\beta}$ for all j gives

$$
\left(\bar{\boldsymbol{Z}}_{1i}-\bar{\boldsymbol{Z}}_{2i}\right) = (\boldsymbol{Z}_{1i}-\boldsymbol{Z}_{2i}) + \sum_{j=1}^{n_i}\left[O(\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)-\psi(z_{ij})) + O(\frac{\partial\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)}{\partial\boldsymbol{\beta}} - \frac{\partial\tilde{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)}{\partial\boldsymbol{\beta}})\right].
$$

According to (3.61), $\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)-\psi(z_{ij}) = O_p(1/\sqrt{K}) + O(b^2)$. Here we require that $b \propto K^{-\alpha}$ satisfies $\alpha > 1/4$ to ensure the $\sqrt{K}-$consistency of nonparametric function estimator $\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)$. Then $\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2) - \psi(z_{ij}) = o_p(1)$. Next, by the law of large numbers for independent random variables (Breiman (1968, Theorem 3.27)), $\frac{\partial\hat{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)}{\partial\boldsymbol{\beta}} - \frac{\partial\tilde{\psi}(z_{ij};\boldsymbol{\beta},\sigma_\tau^2)}{\partial\boldsymbol{\beta}} = o_p(1)$. So by neglecting the higher order terms, one may obtain

$$\left(\bar{\boldsymbol{Z}}_{1i} - \bar{\boldsymbol{Z}}_{2i}\right) \approx \left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right). \tag{3.69}$$

We now turn back to Lemma 3.8 and derive the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ as in the following theorem.

Theorem 3.1. *The SGQL estimator $\hat{\boldsymbol{\beta}}$ (the solution of (3.20)) has the limiting (as $K \to \infty$) multivariate normal distribution given as*

$$\sqrt{K}\left\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} - O(b^2)\right\} \quad \xrightarrow{D} \quad N_p(0, \mathbf{V}_{\boldsymbol{\beta}}), \tag{3.70}$$

where

$$\mathbf{V}_{\boldsymbol{\beta}} = \left[\boldsymbol{F}(\boldsymbol{\beta})\right]^{-1} \frac{1}{K}\left[\sum_{i=1}^{K}\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)\boldsymbol{\Sigma}_i\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)^{\top}\right]\left[\boldsymbol{F}(\boldsymbol{\beta})\right]^{-1}.$$

Proof. By using (3.65) in (3.62), one obtains

$$\sqrt{K}\left\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} - O(b^2)\right\} \quad \approx \quad \left[\boldsymbol{F}(\boldsymbol{\beta})\right]^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left(\bar{\boldsymbol{Z}}_{1i} - \bar{\boldsymbol{Z}}_{2i}\right)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right). \tag{3.71}$$

Now because $\bar{\boldsymbol{Z}}_{1i} - \bar{\boldsymbol{Z}}_{2i}$ can be treated as semi-parametric longitudinal (covariance) weight matrix which we have approximated as in (3.69), we rewrite the approximate equation (3.71) as

$$\sqrt{K}\left\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} - O(b^2)\right\} \quad \approx \quad \left[\boldsymbol{F}(\boldsymbol{\beta})\right]^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right). \tag{3.72}$$

Next, define

$$\bar{\boldsymbol{f}}_K = \frac{1}{K}\sum_{i=1}^{K}\boldsymbol{f}_i = \frac{1}{K}\sum_{i=1}^{K}\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right), \tag{3.73}$$

where $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_K$ are independent of each other as they are collected from K independent individuals. However, they are not identically distributed because

$$\boldsymbol{Y}_i \sim \left[\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right], \tag{3.74}$$

where the mean vectors and covariance matrices are different for different individuals. By (3.74), it follows from (3.73) that

$$\begin{aligned}
\mathrm{E}[\bar{\boldsymbol{f}}_K] &= 0 \\
\mathrm{cov}[\bar{\boldsymbol{f}}_K] &= \frac{1}{K^2}\sum_{i=1}^{K}\mathrm{cov}[\boldsymbol{f}_i] \\
&= \frac{1}{K^2}\sum_{i=1}^{K}\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)\boldsymbol{\Sigma}_i\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)^\top \\
&= \frac{1}{K^2}\boldsymbol{V}_K^*. \qquad (3.75)
\end{aligned}$$

If the multivariate version of Lindeberg's condition holds, that is,

$$\lim_{K\to\infty}\boldsymbol{V}_K^{*\,-1}\sum_{i=1}^{K}\sum_{(\boldsymbol{f}_i^\top\boldsymbol{V}_K^{*\,-1}\boldsymbol{f}_i)>\epsilon}\boldsymbol{f}_i\boldsymbol{f}_i^\top g(\boldsymbol{f}_i) = 0 \tag{3.76}$$

for all $\epsilon > 0$, $g(\cdot)$ being the probability distribution of $\boldsymbol{f}_i$, then the Lindeberg-Feller central limit theorem (Amemiya, 1985, Theorem 3.3.6; McDonald, 2005, Theorem 2.2) implies that

$$K(\boldsymbol{V}_K^*)^{-\frac{1}{2}}\bar{\boldsymbol{f}}_K \xrightarrow{D} N_p(0, \boldsymbol{I}_p). \tag{3.77}$$

Then (3.72) gives

$$\sqrt{K}\left\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} - O(b^2)\right\} = [\boldsymbol{F}(\boldsymbol{\beta})]^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + o_p(1)$$
$$= [\boldsymbol{F}(\boldsymbol{\beta})]^{-1}\sqrt{K}\,\bar{\boldsymbol{f}}_K + o_p(1)$$
$$\xrightarrow{D} N_p(0, K\frac{1}{K^2}\boldsymbol{F}^{-1}\boldsymbol{V}_K^*\boldsymbol{F}^{-1}) = N_p(0, \boldsymbol{V}_{\boldsymbol{\beta}}), \tag{3.78}$$

yielding the Theorem. □

Note that because $b \propto K^{-\alpha}$, for $\sqrt{K}$-consistency of $\hat{\boldsymbol{\beta}}$, we need to have $Kb^4 \to 0$ as $K \to \infty$, which happens when $1/4 < \alpha \leq 1/3$ (see, for example, Lin and Carroll (2001) for upper limit).

3.3.3 Consistency of the SGQL estimator of σ_τ^2 and its asymptotic normal distribution

Notice that the SGQL estimating equation (3.20) for $\boldsymbol{\beta}$ has the form

$$\sum_{i=1}^{K}\frac{\partial\bar{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))}{\partial\boldsymbol{\beta}}\bar{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))\left(\boldsymbol{y}_i - \bar{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))\right) = 0,$$

whereas the SGQL estimating equation (3.33) for σ_τ^2 has a similar but different form given by

$$\sum_{i=1}^{K}\bar{\boldsymbol{Q}}_{1i}\left(\boldsymbol{u}_i - \bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))\right) = 0, \tag{3.79}$$

where

$$\bar{\boldsymbol{Q}}_{1i} = \frac{\partial\bar{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))}{\partial\sigma_\tau^2}\bar{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)).$$

One of the big differences between these estimating equations lies in the fact that the nonparametric function estimate $\hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)$ involved in both equations is a function of the first order response $\{y_{ij}\}$ (3.15), while $\boldsymbol{y}_i$ is the vector of first order responses in the estimating equation for $\boldsymbol{\beta}$, but $\boldsymbol{u}_i$ involved in the estimating equation for σ_τ^2 is a vector of squared responses. This difference will come to play when we derive the asymptotic properties of the SGQL estimator of σ_τ^2, say $\hat{\sigma}_\tau^2$. In preparation for the main result given in Theorem 3.2, we first prove the following two lemmas.

Lemma 3.10. *We express the estimating equation (3.79) as* $KM_K = 0$, *where*

$$M_K = \frac{1}{K} \sum_{i=1}^{K} \bar{\boldsymbol{Q}}_{1i} \left(\boldsymbol{u}_i - \bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)) \right).$$

Then we can write

$$\hat{\sigma}_\tau^2 - \sigma_\tau^2 = L^{-1} M_K + o_p(1/\sqrt{K}), \tag{3.80}$$

where

$$L = \mathrm{E}\left[\frac{\partial \bar{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))}{\partial \sigma_\tau^2} \bar{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2)) \frac{\partial \bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\boldsymbol{\beta}, \sigma_\tau^2))}{\partial \sigma_\tau^2} \right].$$

Proof. Because the SGQL estimator $\hat{\sigma}_\tau^2$ of σ_τ^2 obtained from (3.79) [see also (3.33)] satisfies $M_K(\hat{\sigma}_\tau^2) = 0$, a linear Taylor series expansion, similar to (3.63), about σ_τ^2 provides

$$\begin{aligned} \hat{\sigma}_\tau^2 - \sigma_\tau^2 &= L_K^{-1} M_K + O(|\hat{\sigma}_\tau^2 - \sigma_\tau^2|^2) \\ &= L_K^{-1} M_K + o_p(1/\sqrt{K}), \end{aligned} \tag{3.81}$$

where $L_K = \partial M_K / \partial \sigma_\tau^2$ and it has the formula given by

$$\begin{aligned} L_K &= \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \bar{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))}{\partial \sigma_\tau^2}\, \bar{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \rho, \hat{\psi}(\cdot)) \frac{\partial \bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(\cdot))}{\partial \sigma_\tau^2} \\ &= \frac{1}{K}\sum_{i=1}^{K} L_i^*. \end{aligned} \tag{3.82}$$

Here we in fact applied Lemma 3.11 and $b \propto K^{-\alpha}$ with $1/5 \le \alpha \le 1/3$ (Lin and Carroll, 2001) to learn that $O(|\hat{\sigma}_\tau^2 - \sigma_\tau^2|^2)$ is in the order of $o_p(1/\sqrt{K})$. Note that $\mathrm{E}(L_i^*) = L$. Then it is easy to prove that $L_K = L + O_p(1/\sqrt{K})$ as $K \to \infty$, and the lemma is proven. □

Lemma 3.11. *M_K in Lemma 3.10 may be shown to satisfy*

$$M_K = \frac{1}{K}\sum_{i=1}^{K} \bar{\boldsymbol{Q}}_{1i}\left[\boldsymbol{u}_i - \boldsymbol{\lambda}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right] - \frac{1}{K}\sum_{i=1}^{K} \bar{\boldsymbol{Q}}_{2i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + O(b^2) + o_p(1/\sqrt{K}), \tag{3.83}$$

where $\bar{\boldsymbol{Q}}_{1i}$ is defined in (3.79), and $\bar{\boldsymbol{Q}}_{2i} = \left(\bar{Q}_{2i1}, \cdots, \bar{Q}_{2ij}, \cdots, \bar{Q}_{2in_i}\right)^\top$ with

$$\bar{Q}_{2ij} = \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}}\sum_{k'=1}^{n_{i'}} \frac{1}{B_K(z_{i'k'})} W_{i'j'k'}^* \frac{p_{ij}(z_{i'k'})}{1 + \mu_{ij}(z_{i'k'})\,(e^{\sigma_\tau^2} - 1)}, \tag{3.84}$$

where

$$W_{ijk}^* = \frac{\partial \bar{\lambda}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \sigma_\tau^2}\, v_{2i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho) \frac{\partial \bar{\lambda}_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik}))}{\partial \psi(z_{ik})} \tag{3.85}$$

with $v_{2i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)$ being the (j,k)th element of the $n_i \times n_i$ inverse fourth order moments matrix $\bar{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)$.

Proof. Write M_K from (3.79) as

$$M_K = M_{1K} - M_{2K}, \tag{3.86}$$

where

$$\begin{aligned} M_{1K} &= \frac{1}{K}\sum_{i=1}^{K} \bar{\boldsymbol{Q}}_{1i}\left[\boldsymbol{u}_i - \boldsymbol{\lambda}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right] \\ M_{2K} &= \frac{1}{K}\sum_{i=1}^{K} \bar{\boldsymbol{Q}}_{1i}\left[\bar{\boldsymbol{\lambda}}_i(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\boldsymbol{\psi}}(\boldsymbol{z}_i; \boldsymbol{\beta}, \sigma_\tau^2)) - \boldsymbol{\lambda}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))\right]. \end{aligned} \tag{3.87}$$

Now because

$$\bar{\boldsymbol{\lambda}}_i(\cdot) = [\bar{\lambda}_{i1}(\cdot), \ldots, \bar{\lambda}_{ij}(\cdot), \ldots, \bar{\lambda}_{in_i}(\cdot)]^\top, \text{ and } \boldsymbol{\lambda}_i(\cdot) = [\lambda_{i1}(\cdot), \ldots, \lambda_{ij}(\cdot), \ldots, \lambda_{in_i}(\cdot)]^\top,$$

using the notation for $\bar{\boldsymbol{Q}}_{1i}$ from (3.79), we express M_{2K} in (3.87) as

$$\begin{aligned} M_{2K} &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} \frac{\partial \bar{\lambda}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2))}{\partial \sigma_\tau^2}\, v_{2i}^{jk}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}, \rho)\left[\bar{\lambda}_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ik}; \boldsymbol{\beta}, \sigma_\tau^2))\right. \\ &- \left.\lambda_{ik}(\boldsymbol{\beta}, \sigma_\tau^2, \psi(z_{ik}))\right]. \end{aligned} \tag{3.88}$$

A Taylor expansion of $\bar{\lambda}_{ij}(\boldsymbol{\beta}, \sigma_\tau^2, \hat{\psi}(z_{ij}; \boldsymbol{\beta}, \sigma_\tau^2))$ with respect to $\psi(z_{ij})$ for all $i = 1, \ldots, K$ and $j = 1, \ldots, n_i$, reduces M_{2K} in (3.88) to

$$\begin{aligned} M_{2K} &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i}\left\{W_{ijk}^*\left[\hat{\psi}(z_{ik}; \boldsymbol{\beta}, \sigma_\tau^2) - \psi(z_{ik})\right]\right. \\ &+ \left. O_p\left(\left[\hat{\psi}(z_{ik}; \boldsymbol{\beta}, \sigma_\tau^2) - \psi(z_{ik})\right]^2\right)\right\}, \end{aligned} \tag{3.89}$$

where W_{ijk}^* is defined in (3.85). Further by using the formula for $\hat{\psi}(z_{ik}; \boldsymbol{\beta}, \sigma_\tau^2) - \psi(z_{ik})$

from (3.61), M_{2K} in (3.89) may be re-expressed as

$$M_{2K} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i}\left\{W^*_{ijk}\left[A_K(z_{ik}) + O(b^2)\right] + O_p\left(\left[O_p(1/\sqrt{K}) + O(b^2)\right]^2\right)\right\}$$
$$= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} W^*_{ijk}A_K(z_{ik}) + O(b^2) + o_p(1/\sqrt{K}), \quad (3.90)$$

where A_K is given by (3.56). Then by using the formula for A_K for (3.56), one may obtain

$$M_{2K} = \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}}\sum_{k'=1}^{n_{i'}} W^*_{i'j'k'}\left\{\frac{1}{B_K(z_{i'k'})}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_{i'k'})\,\frac{Y_{ij}-\mu_{ij}}{1+\mu_{ij}(z_{i'k'})\left(\exp(\sigma^2_\tau)-1\right)}\right\}$$
$$+ O(b^2) + o_p(1/\sqrt{K})$$
$$= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left\{\frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}}\sum_{k'=1}^{n_{i'}}\frac{1}{B_K(z_{i'k'})}W^*_{i'j'k'}\,\frac{p_{ij}(z_{i'k'})}{1+\mu_{ij}(z_{i'k'})\left(\exp(\sigma^2_\tau)-1\right)}\right\}(Y_{ij}-\mu_{ij})$$
$$+ O(b^2) + o_p(1/\sqrt{K})$$
$$= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\bar{Q}_{2ij}\,(Y_{ij}-\mu_{ij}) + O(b^2) + o_p(1/\sqrt{K})$$
$$= \frac{1}{K}\sum_{i=1}^{K}\bar{\boldsymbol{Q}}_{2i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + O(b^2) + o_p(1/\sqrt{K}). \quad (3.91)$$

The above results for M_{1K} (3.87) and M_{2K} (3.91) together yield the lemma. $\square$

Let $\boldsymbol{Q}_{1i}$ and $\boldsymbol{Q}_{2i}$ be the $\bar{\boldsymbol{Q}}_{1i}$ and $\bar{\boldsymbol{Q}}_{2i}$, respectively, with $\hat{\psi}(z_{il};\boldsymbol{\beta},\sigma^2_\tau)$ being replaced by its true value $\psi(z_{il})$, and $\partial\hat{\psi}(z_{il};\boldsymbol{\beta},\sigma^2_\tau)/\partial\sigma^2_\tau$ given in (3.37) being replaced by

$$\frac{\partial\tilde{\psi}(z_{il};\boldsymbol{\beta},\sigma^2_\tau)}{\partial\sigma^2_\tau} = -\frac{1}{2} - \exp(\sigma^2_\tau)\left[\frac{\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_{il})\mathrm{E}\left[\left\{\frac{y_{ij}-\bar{\mu}_{ij}(z_{il})}{[1+\bar{\mu}_{ij}(z_{il})(\exp(\sigma^2_\tau)-1)]^2}\right\}\bar{\mu}_{ij}(z_{il})\right]}{\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_{il})\mathrm{E}\left[\left\{\frac{1+y_{ij}(\exp(\sigma^2_\tau)-1)}{[1+\bar{\mu}_{ij}(z_{il})(\exp(\sigma^2_\tau)-1)]^2}\right\}\bar{\mu}_{ij}(z_{il})\right]}\right] \quad (3.92)$$

for all l. Then with a similar argument as that for (3.69), one may obtain

$$\bar{\boldsymbol{Q}}_{1i} \approx \boldsymbol{Q}_{1i} \quad \text{and} \quad \bar{\boldsymbol{Q}}_{2i} \approx \boldsymbol{Q}_{2i}. \tag{3.93}$$

We now turn back to (3.80) and derive the asymptotic distribution of $\hat{\sigma}_\tau^2$ as in the following theorem.

Theorem 3.2. *The SGQL estimator $\hat{\sigma}_\tau^2$ (the solution of (3.33)) has the limiting (as $K \to \infty$) normal distribution given as*

$$\sqrt{K}\left\{\hat{\sigma}_\tau^2 - \sigma_\tau^2 - O\left(b^2\right)\right\} \overset{D}{\to} N(0, V_{\sigma_\tau^2}) \quad \text{as } K \to \infty, \tag{3.94}$$

where

$$V_{\sigma_\tau^2} = L^{-1}\frac{1}{K}\sum_{i=1}^{K}\left[\boldsymbol{Q}_{1i}\boldsymbol{\Omega}_i\boldsymbol{Q}_{1i}^T + \boldsymbol{Q}_{2i}\boldsymbol{\Sigma}_i\boldsymbol{Q}_{2i}^T - 2\boldsymbol{Q}_{1i}\mathrm{Cov}(\boldsymbol{U}_i, \boldsymbol{Y}_i)\boldsymbol{Q}_{2i}^T\right] L^{-1}. \tag{3.95}$$

Proof. By applying Lemma 3.11, it now follows from Lemma 3.10 that

$$\begin{aligned}\sqrt{K}\left\{\hat{\sigma}_\tau^2 - \sigma_\tau^2 - O(b^2)\right\} \approx\ & L^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\bar{\boldsymbol{Q}}_{1i}[\boldsymbol{u}_i - \boldsymbol{\lambda}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))] \\ & - L^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\bar{\boldsymbol{Q}}_{2i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i). \end{aligned}\tag{3.96}$$

Then similar to (3.72), by using (3.93), we can rewritten (3.96) as

$$\begin{aligned}\sqrt{K}\left\{\hat{\sigma}_\tau^2 - \sigma_\tau^2 - O(b^2)\right\} \approx\ & L^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\boldsymbol{Q}_{1i}[\boldsymbol{u}_i - \boldsymbol{\lambda}_i(\boldsymbol{\beta}, \sigma_\tau^2, \boldsymbol{\psi}(\boldsymbol{z}_i))] \\ & - L^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\boldsymbol{Q}_{2i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i). \end{aligned}\tag{3.97}$$

Further by using similar arguments as in the proof of Theorem 3.1, one may apply the

Lindeberg-Feller central limit theorem for non-identically distributed random variables (Amemiya, 1985, Theorem 3.3.6), and proves the theorem. $\square$

Similar to the condition for the $\sqrt{K}$-consistency of $\hat{\boldsymbol{\beta}}$, for $\sqrt{K}$-consistency of $\hat{\sigma}_\tau^2$, we need to have $O(b^2) = Kb^4 \to 0$ as $K \to \infty$, that is, $1/4 < \alpha \leq 1/3$ (see Lin and Carroll (2001), for example, for the upper limit).

Note that in this section we have derived the asymptotic properties of $\hat{\sigma}_\tau^2$ which is obtained by applying the SGQL estimation approach using squared responses as discussed in Section 3.3.1. The derivation of the asymptotic properties for the estimator of σ_τ^2 obtained by applying the variation of the SGQL approach from Section 3.3.2 or 3.3.3 will be similar, and hence omitted for the interest of space.

3.3.4 Consistency of the moment estimator of ρ

The consistency of the moment estimator $\hat{\rho}$ obtained in Section 3.2.4 is given by the following lemma:

Lemma 3.12. *For a_1, b_1 and g_1 defined by (3.51), (3.53) and (3.54) respectively, the moment estimator $\hat{\rho} = \frac{a_1 - b_1}{g_1}$ obtained from (3.52) is a consistent estimator for the longitudinal correlation index parameter ρ.*

Proof. Recall from (3.49) that $Y_{ij}^* = (Y_{ij} - \mu_{ij}) / (\sigma_{ijj})^{1/2}$. It is obvious that

$$\begin{aligned} \mathrm{E}\left(Y_{ij}^{*2}\right) &= 1 \quad \text{for all } i \text{ and } j \qquad (3.98) \\ \Rightarrow \quad \mathrm{E}\left[\sum_{j=1}^{n_i}\left(Y_{ij}^{*2} - 1\right)\right] &= 0 \quad \text{for all } i = 1, \ldots, K \end{aligned}$$

and

$$\mathrm{E}\left[\left(\frac{Y_{ij} - \mu_{ij}}{\sqrt{\sigma_{ijj}}}\right)\left(\frac{Y_{i,j+1} - \mu_{i,j+1}}{\sqrt{\sigma_{i,j+1,j+1}}}\right)\right] = \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}} \qquad (3.99)$$

$$\Rightarrow \quad \mathrm{E}\left[\sum_{j=1}^{n_i-1}\left(Y_{ij}^*Y_{i,j+1}^* - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right)\right] \quad = \quad 0 \quad \text{for all } i = 1, \ldots, K.$$

Now because $\mathrm{E}\left[\left(\sum_{j=1}^{n_i-1}\left[Y_{ij}^*Y_{i,j+1}^* - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]\right)^2\right]$ by (3.99) and $\mathrm{E}\left[\left(\sum_{j=1}^{n_i}\left[Y_{ij}^{*2} - 1\right]\right)^2\right]$ by (3.98) are all functions of μ_{ij}, σ_τ^2 and ρ, they are bounded under the assumption that μ_{ij} and n_i are all bounded. Thus for a sufficiently large but finite m_0, one may write

$$\begin{aligned} \mathrm{E}\left[\left(\sum_{j=1}^{n_i-1}\left[Y_{ij}^*Y_{i,j+1}^* - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]\right)^2\right] &< m_0, \\ \text{and also } \mathrm{E}\left[\left(\sum_{j=1}^{n_i}\left[Y_{ij}^{*2} - 1\right]\right)^2\right] &< m_0, \end{aligned} \tag{3.100}$$

for all $i = 1, \ldots, K$. Now because Y_{ij}'s are independent for different i, it follows from the law of large numbers for independent random variables [Breiman (1968, Theorem 3.27)] that

$$\begin{aligned} & \frac{\sum_{i=1}^K\left[\sum_{j=1}^{n_i-1}\left(y_{ij}^*y_{i,j+1}^* - \frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right)\right]}{\sum_{i=1}^K(n_i-1)} \xrightarrow{P} 0 \\ & \Rightarrow \frac{\sum_{i=1}^K\left[\sum_{j=1}^{n_i-1} y_{ij}^*y_{i,j+1}^*\right]}{\sum_{i=1}^K(n_i-1)} = \frac{\rho\sum_{i=1}^K\left[\sum_{j=1}^{n_i-1}\frac{\mu_{ij}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]}{\sum_{i=1}^K(n_i-1)} \\ + \quad & \frac{(\exp(\sigma_\tau^2)-1)\sum_{i=1}^K\left[\sum_{j=1}^{n_i-1}\frac{\mu_{ij}\mu_{i,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]}{\sum_{i=1}^K(n_i-1)} + o_p(1), \end{aligned} \tag{3.101}$$

and

$$\frac{\sum_{i=1}^K\left[\sum_{j=1}^{n_i}\left({y_{ij}^*}^2 - 1\right)\right]}{\sum_{i=1}^K n_i} \quad \xrightarrow{P} \quad 0 \qquad \Rightarrow$$

$$\frac{\sum_{i=1}^{K}\left[\sum_{j=1}^{n_i} {y_{ij}^*}^2\right]}{\sum_{i=1}^{K} n_i} \quad = \quad 1 + o_p(1). \tag{3.102}$$

Next, dividing (3.101) by (3.102) and using the notations a_1, b_1 and g_1 from (3.51)−(3.54), we can write

$$a_1(1 + o_p(1)) = \rho\, g_1 + b_1 + o_p(1) \Rightarrow a_1 + a_1\, o_p(1) = \rho\, g_1 + b_1 + o_p(1). \tag{3.103}$$

Here

$$a_1\, o_p(1) = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1} y_{ij}^* y_{i,j+1}^* / \sum_{i=1}^{K}(n_i - 1)}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} {y_{ij}^*}^2 / \sum_{i=1}^{K} n_i}\, o_p(1) = o_p(1).$$

This is because the numerator and denominator for a_1 are finite by the law of large numbers (Breiman, 1968, Theorem 3.27). Consequently, from (3.103) we obtain

$$a_1 + o_p(1) = \rho\, g_1 + b_1 + o_p(1) \quad \Rightarrow \quad \hat{\rho} = \frac{a_1 - b_1}{g_1} = \rho + o_p(1),$$

or equivalently,

$$\hat{\rho} = \frac{a_1 - b_1}{g_1} \quad \xrightarrow{P} \quad \rho \quad \text{as } K \to \infty. \tag{3.104}$$

Hence, the lemma follows. $\square$

We remark that the consistency result in (3.104) remains valid when $\psi(\cdot)$ in μ_{ij}'s is replaced by its consistent estimate $\hat{\psi}(\cdot)$.

3.4 A simulation study

The objective of our simulation study in this section is to examine the finite sample performance of the (1) SQL approach for $\psi(\cdot)$ estimation; (2) SGQL estimation for $\boldsymbol{\beta}$ and σ_τ^2; and (3) SMM estimation for the correlation index parameter ρ.

3.4.1 Design construction

For the purpose, we select the parameters, primary and secondary covariates, and the nonparametric function as follows:

1. Parameters selection: We consider the following four sets of parameter values.

Set 1: $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 0.5$, $\rho = 0.5$;

Set 2: $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 0.5$, $\rho = 0.8$;

Set 3: $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 1.0$, $\rho = 0.5$; and

Set 4: $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) = (0.5, 0.5)$, $\sigma_\tau^2 = 1.0$, $\rho = 0.8$.

2. Primary covariate selection: For the primary covariate selection, we choose $n_i = 4$ equi-spaced time points for all $i = 1, \ldots, K$, with $K = 100$. Next, because $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ is the effect of two time dependent primary covariates, we choose these covariates as

$$x_{ij1}(j) = \begin{cases} \frac{1}{2} & \text{for } i = 1, \ldots, 25 \text{ and } j = 1, 2 \\ 1 & \text{for } i = 1, \ldots, 25 \text{ and } j = 3, 4 \\ \frac{-1}{2} & \text{for } i = 26, \ldots, 75 \text{ and } j = 1 \\ 0 & \text{for } i = 26, \ldots, 75 \text{ and } j = 2, 3 \\ \frac{1}{2} & \text{for } i = 26, \ldots, 75 \text{ and } j = 4 \\ \frac{j}{2n_i} & \text{for } i = 76, \ldots, 100 \text{ and } j = 1, 2, 3, 4 \end{cases}$$

$$x_{ij2}(j) = \begin{cases} \frac{j-2.5}{2n_i} & \text{for } i = 1, \ldots, 50 \text{ and } j = 1, 2, 3, 4 \\ 0 & \text{for } i = 51, \ldots, 100 \text{ and } j = 1, 2 \\ \frac{1}{2} & \text{for } i = 51, \ldots, 100 \text{ and } j = 3, 4. \end{cases} \tag{3.105}$$

Note that these covariate values are the same as in Sutradhar (2010, p. 188). These values are chosen to reflect the variable time dependence for the different groups of individuals. Thus, the choice is quite general. One may choose other specific covariates depending on the situations.

3. Random effects generation: The random effects τ_i for $i = 1, \ldots, 100$, are generated from $N(0, 1)$ distribution.

4. Secondary covariate selection: For a given i ($i = 1, \ldots, 100$), we choose a value for z_{ij} from a uniform (U) distribution, namely

$$z_{ij} \sim U[j - 0.5, j + 0.5], \tag{3.106}$$

for $j = 1, \ldots, n_i = 4$. Note that for each $j = 1, \ldots, 4$, the interval $[j - 0.5, j + 0.5]$ was divided into 25 (alternatively it could be 50 or 100, and so on) equi-spaced points allowing one value to be chosen from 25 values. Thus, altogether $n_i = 4$ values were chosen from j-related four intervals. This was independently repeated for $K = 100$ individuals. Consequently, these 400 values are expected to be dense and they reflect the time dependence.

5. Nonparametric function selection: We chose, for example, a quadratic non-parametric function given by

$$\psi(z_{ij}) = 0.3 + 0.2\left(z_{ij} - \frac{n_i + 1}{2}\right) + 0.05\left(z_{ij} - \frac{n_i + 1}{2}\right)^2 \tag{3.107}$$

with $n_i = 4$, where z_{ij} is generated by (3.106). Remark that in practice this nonparametric function influencing y_{ij} would be unknown.

3.4.2 Data generation

We use the design selected from the last section into the SGLMM (or SMM in brief, (3.3)−(3.4)) and generate the data. More specifically, we use (3.3) and (3.4) to generate repeated Poisson observations $\{y_{ij}, j = 1, \ldots, n_i; i = 1, \ldots, K\}$ conditional on the random effects and nonparametric function, where random effects and nonparametric function are chosen as indicated above.

3.4.3 Naive estimation (ignoring $\psi(\cdot)$) effect on $\boldsymbol{\beta}$, σ_τ^2 and ρ estimates

When repeated Poisson count data are generated under the present SGLMM (3.3)−(3.4) following Sections 3.4.1 and 3.4.2, but one ignores the presence of $\psi(\cdot)$ in the model and makes an attempt to estimate the parameters ($\boldsymbol{\beta}, \sigma_\tau^2$ and ρ) by treating the data as though they were generated from the GLMM ($\psi(\cdot) = 0$), the estimates will be biased. To have an idea of the magnitude of bias, we examine the performance of such naive GQL (NGQL) estimators by repeating the data generation 1000 times and computing the simulated mean (SM), simulated standard error (SSE), and simulated mean squared error (SMSE) of the NGQL estimates for $\boldsymbol{\beta}$ and σ_τ^2, and moment estimate of ρ. The parameter values and their simulated estimates are shown in Table 3.1.

True $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$	σ^2_τ	ρ	Quantity	$\hat{\beta}_{1,NGQL}$	$\hat{\beta}_{2,NGQL}$	$\hat{\sigma}^2_{\tau,NGQL}$	$\hat{\rho}_{Moment}$
$\boldsymbol{\beta} = (0.5, 0.5)^\top$	0.5	0.5	SM	0.9483	1.1209	0.6829	0.1850
			SSE	0.1199	0.1778	0.1468	0.1640
			MSE	0.2153	0.4171	0.0550	
		0.8	SM	0.9638	1.1134	0.6836	0.5261
			SSE	0.0995	0.1550	0.1445	0.1540
			MSE	0.2250	0.4003	0.0545	
	1.0	0.5	SM	0.9669	1.0962	1.1364	0.1100
			SSE	0.1227	0.1751	0.2656	0.1662
			MSE	0.2331	0.3861	0.0890	
		0.8	SM	0.9704	1.0957	1.1461	0.3337
			SSE	0.1031	0.1469	0.2960	0.2288
			MSE	0.2319	0.3764	0.1088	

Table 3.1: Simulated means (SMs), simulated standard errors (SSEs) and mean squared errors (MSEs) of NGQL estimates (ignoring the presence of $\psi(\cdot)$) of regression parameters $\boldsymbol{\beta}$ and random effects variance σ^2_τ under non-stationary AR(1) correlation model (3.3) and (3.4) for selected values of correlation index parameter ρ with $K = 100$, $n_i = 4$; based on 1000 simulations.

As expected, the results in Table 3.1 show that the estimates of $\boldsymbol{\beta}$ and σ^2_τ are highly biased. For example, when $\rho = 0.5$, for the true regression parameter $\boldsymbol{\beta} = (0.5, 0.5)^\top$ and random effects variance $\sigma^2_\tau = 0.5$, the estimated values of $\boldsymbol{\beta}$ and σ^2_τ are found to be $(0.9483, 1.1209)^\top$ and 0.6829, respectively. The estimate for $\rho = 0.5$ was found to be 0.185. Clearly all of these naive estimates computed by ignoring $\psi(z_{ij})$ are useless, and hence one must take $\psi(z_{ij})$ into account in estimating these regression, overdispersion and correlation index parameters. This will require the consistent estimation of the nonparametric function as well, which was discussed in Section 3.2.1.

3.4.4 Main simulation results

We now examine the performance of the proposed semi-parametric estimation approach discussed in Section 3.2 for the estimation of the function $\psi(z_{ij})$, and all the parameters ($\boldsymbol{\beta}, \sigma_\tau^2$ and ρ). The overdispersion parameter σ_τ^2 was estimated by using squared response based exact (SR-exact), corrected squared response based exact (CSR-exact), and corrected squared normal (CSR-normal) techniques as discussed in Section 3.2.3. We also examine the performance of the SGQL approach by pretending that the correlation index parameter ρ is zero. Recall that in the present setup, $\rho = 0$ does not mean the repeated responses are independent. The independence requires that both $\rho = 0$ and $\sigma_\tau^2 = 0$. All estimates (simulated mean, SM) along with their standard errors (SSE) and mean square errors (MSE) are obtained based on 1000 simulations. The results are provided in Table 3.2 for $\boldsymbol{\beta}$, σ_τ^2 and ρ parameters. The SQL estimate for $\psi(\cdot)$ is displayed in Fig. 3.1. Note that this estimate uses σ_τ^2 estimated by the exact weight matrix discussed above. Here the bandwidth in $\psi(\cdot)$ estimation is chosen as $b = K^{-1/5}$ (Pagan and Ullah, 1999, Altman, 1990, Horowitz, 2009) to minimize bias and variance of the nonparametric function estimates, instead of considering consistency of the estimators.

Figure 3.1 shows that the SQL approach estimates the true nonparametric curve well. The estimated curve almost coincides with the true curve when overdispersion index parameter is small , that is, $\sigma_\tau^2 = 0.5$. This holds for small and large correlation index parameter (ρ) values. The curve estimate is less satisfactory when $\sigma_\tau^2 = 1.0$. This happens because $\sigma_\tau^2 = 1.0$ produces large overdispersion in the data and, as the results of Table 3.2 show, the estimates $\hat{\sigma}_\tau^2$ are slightly biased when $\sigma_\tau^2 = 1.0$.

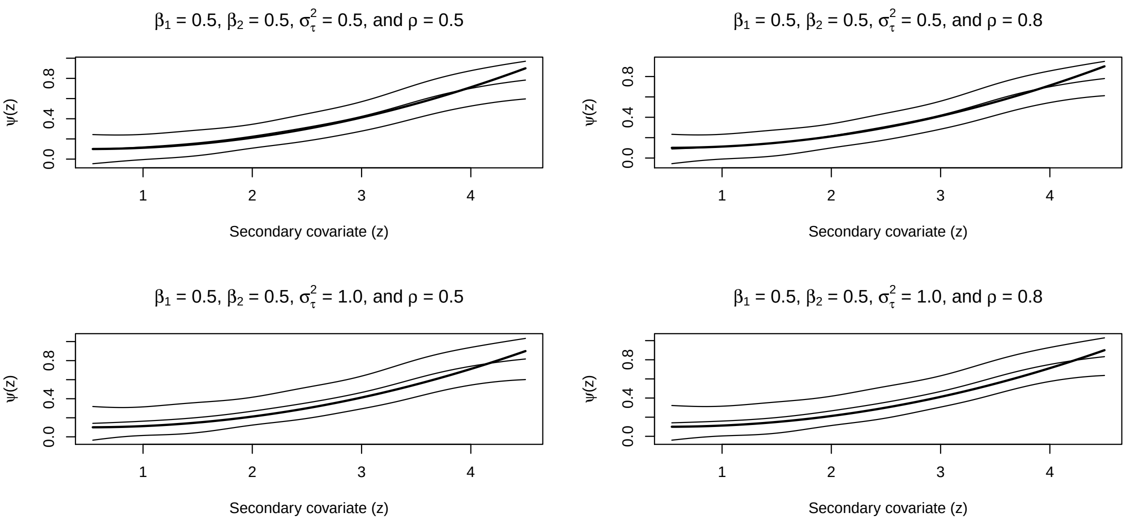

Figure 3.1: The plot for $\psi(\cdot)$ estimation for the approach with σ_τ^2 estimated by the exact weight matrix given in Section 3.2.3.1. The thick curve is the true $\psi(\cdot)$ function value. The thinner curves are the estimated $\psi(\cdot)$ value and the one standard error curves. The bandwidth $b = K^{-1/5}$.

Next, the results from Table 3.2 indicate that the main regression parameters $\beta_1 = \beta_2 = 0.5$ are estimated very well by the proposed SGQL approach irrespective of the SGQL procedures (SR-Approx ($\rho = 0$), SR-exact, CSR-exact or CSR-normal) used for the estimation of σ_τ^2. This estimation pattern holds whether correlation index ρ is small (0.5) or large (0.8). For example, for large $\rho = 0.8$ and small $\sigma_\tau^2 = 0.5$ (estimated by SR-exact approach), the SGQL estimates of $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (0.5, 0.5)^\top$ are $(0.4940, 0.4844)^\top$ with MSEs $(0.0165, 0.0516)^\top$. The estimates are similar even when σ_τ^2 is large, specifically when $\sigma_\tau^2 = 1.0$, the estimates are $(0.4922, 0.4701)^\top$ with MSEs $(0.0163, 0.0471)^\top$. As far as the estimation of correlation parameters σ_τ^2 and ρ is concerned, the SGQL approaches for σ_τ^2 and the method of moments for ρ, work very well when σ_τ^2 is small. For large $\sigma_\tau^2 = 1.0$, the estimates for both parameters are slightly biased. For example, when $\rho = 0.8$, for $\sigma_\tau^2 = 0.5$, the CSR-normal weight based SGQL approach produces the estimate $\hat{\sigma}_\tau^2 = 0.4739$ with MSE 0.0163, and the method of moments yields ρ estimate as 0.7689 with SSE 0.0870. When $\sigma_\tau^2 = 1.0$, the CSR-normal weight based SGQL approach produces an estimate $\hat{\sigma}_\tau^2 = 0.8841$ with MSE 0.0720, and the method of moments yields ρ estimate as 0.7399 with SSE 0.1446, i.e., the bias is slightly larger than when $\sigma_\tau^2 = 0.5$. Thus, the simulation study suggests that the proposed estimation approaches perform quite well when overdispersion is small and they perform reasonably well when overdispersion is large. However, for the cases with large overdispersion, it might be desirable to develop a suitable bias correction approach.

Table 3.2: Simulated means (SMs), simulated standard errors (SSEs) and mean squared errors (MSEs) of the SGQL estimates of regression parameters $\boldsymbol{\beta}$, and the SR-exact, CSR-exact, CSR-normal and SR-Approx ($\rho = 0$) weight matrix based SGQL estimates for the random effects variance σ_τ^2 under non-stationary AR(1) correlation model (3.3) and (3.4) for selected values of correlation index parameter ρ with $K = 100$, $n_i = 4$; based on 1000 simulations. The function $\psi(\cdot)$ is estimated by SQL approach and ρ is estimated using method of moments in all cases. The bandwidth $b = K^{-1/5} = 0.3981072$.

True $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$	σ_τ^2	ρ	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\tau^2$	$\hat{\rho}$
$\boldsymbol{\beta} = (0.5, 0.5)^\top$	0.5	0.5	SR-exact	SM	0.4947	0.4881	0.4899	0.4594
				SSE	0.1576	0.2907	0.1252	0.1254
				MSE	0.0249	0.0846	0.0158	
			SR-Approx ($\rho = 0$)	SM	0.4947	0.4876	0.4803	0.4710
				SSE	0.1576	0.2909	0.1264	0.1180
				MSE	0.0248	0.0847	0.0164	
			CSR-exact	SM	0.4942	0.4872	0.4737	0.4771
				SSE	0.1576	0.2909	0.1265	0.1221
				MSE	0.0248	0.0847	0.0167	
			CSR-normal	SM	0.4941	0.4868	0.4694	0.4824
				SSE	0.1575	0.2908	0.1249	0.1162
				MSE	0.0248	0.0847	0.0165	
		0.8	SR-exact	SM	0.4940	0.4844	0.5013	0.7503
				SSE	0.1283	0.2267	0.1269	0.0959
				MSE	0.0165	0.0516	0.0161	
			SR-Approx ($\rho = 0$)	SM	0.4941	0.4838	0.4792	0.7684
				SSE	0.1284	0.2269	0.1291	0.0839
				MSE	0.0165	0.0517	0.0171	
			CSR-exact	SM	0.4937	0.4843	0.4784	0.7657
				SSE	0.1281	0.2271	0.1265	0.0918
				MSE	0.0164	0.0517	0.0165	
			CSR-normal	SM	0.4936	0.4841	0.4739	0.7689
				SSE	0.1280	0.2269	0.1250	0.0870
				MSE	0.0164	0.0517	0.0163	

Table 3.2: (Continued)

True $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$	σ_τ^2	ρ	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}_\tau^2$	$\hat{\rho}$
	1.0	0.5	SR-exact	SM	0.5032	0.4896	0.8823	0.4470
				SSE	0.1660	0.2809	0.2263	0.1891
				MSE	0.0275	0.0789	0.0650	
			SR-Approx ($\rho = 0$)	SM	0.5021	0.4898	0.8845	0.4569
				SSE	0.1665	0.2811	0.2346	0.1881
				MSE	0.0277	0.0790	0.0683	
			CSR-exact	SM	0.5017	0.4894	0.8768	0.4595
				SSE	0.1665	0.2814	0.2327	0.1901
				MSE	0.0277	0.0792	0.0693	
			CSR-normal	SM	0.5017	0.4881	0.8743	0.4673
				SSE	0.1662	0.2805	0.2336	0.1852
				MSE	0.0276	0.0787	0.0703	
		0.8	SR-exact	SM	0.4922	0.4701	0.8989	0.7221
				SSE	0.1275	0.2152	0.2461	0.1545
				MSE	0.0163	0.0471	0.0707	
			SR-Approx ($\rho = 0$)	SM	0.4918	0.4686	0.8883	0.7368
				SSE	0.1278	0.2152	0.2402	0.1456
				MSE	0.0164	0.0472	0.0701	
			CSR-exact	SM	0.4921	0.4690	0.8875	0.7331
				SSE	0.1277	0.2152	0.2384	0.1500
				MSE	0.0163	0.0472	0.0694	
			CSR-normal	SM	0.4911	0.4685	0.8841	0.7399
				SSE	0.1280	0.2156	0.2421	0.1446
				MSE	0.0165	0.0474	0.0720	

Chapter 4

Semi-parametric dynamic fixed models for longitudinal binary data

In Chapter 2 we have discussed a semi-parametric dynamic model for the analysis of longitudinal count data with fixed regression effects. However, recent studies (Sutradhar, 2010) show that, except for the stationary cases where covariates are time independent, the correlation structures for non-stationary binary data are, in general, different than those for count data. Thus, special attention is needed to model the correlations for the non-stationary (with time dependent covariates) binary data which does not follow from count data models discussed in Chapter 2. In the parametric setup, we refer to Sutradhar (2011, Chapter 7) for such non-stationary correlation models for longitudinal binary data. Specifically, two models, namely the LDCP (linear dynamic conditional probability) and the BDL (binary dynamic logit) models are discussed. The purpose of this chapter is to generalize these models to the semi-parametric setup. To be specific, we develop a semi-parametric LDCP (SLDCP) model in Section 4.1 and a semi-parametric BDL (SBDL) model in Section 4.2.

4.1 SLDCP (semi-parametric linear dynamic conditional probability) model for longitudinal binary data

Recall from Sutradhar (2011) (see also Zeger et al., 1985) that for the binary responses $\{y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}\}$ the LDCP model is defined as

$$\begin{aligned} & \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}) = \Pr[Y_{ij} = 1|\boldsymbol{x}_{ij}] \\ = \; & \frac{\exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta})}{1 + \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta})} \text{ for } j = 1, \cdots, n_i, \text{ and} \quad (4.1) \\ & \lambda_{i,j|j-1}(\boldsymbol{\beta}, \rho; \boldsymbol{x}_{ij}, \boldsymbol{x}_{i,j-1}) = Pr[Y_{ij} = 1|y_{i,j-1}, \boldsymbol{x}_{ij}, \boldsymbol{x}_{i,j-1}] \\ = \; & \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}) + \rho[y_{i,j-1} - \mu_{i,j-1}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j-1})] \text{ for } j = 2, \ldots, n_i. \quad (4.2) \end{aligned}$$

Similar to the semi-parametric fixed model for count data (Chapter 2), the above LDCP model may be generalized to the semi-parametric setup as

$$\begin{aligned} & \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) = \Pr[Y_{ij} = 1|\boldsymbol{x}_{ij}, z_{ij}] \\ = \; & \frac{\exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij}))}{1 + \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij}))} \text{ for } j = 1, \ldots, n_i, \text{ and} \quad (4.3) \\ & \lambda_{i,j|j-1}(\boldsymbol{\beta}, \rho, \psi(\cdot); \boldsymbol{x}_{ij}, \boldsymbol{x}_{i,j-1}, z_{ij}, z_{i,j-1}) \\ = \; & Pr[Y_{ij} = 1|y_{i,j-1}, \boldsymbol{x}_{ij}, \boldsymbol{x}_{i,j-1}, z_{ij}, z_{i,j-1}] \\ = \; & \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) + \rho[y_{i,j-1} - \mu_{i,j-1}(\boldsymbol{\beta}, \boldsymbol{x}_{i,j-1}, \psi(z_{i,j-1}))] \text{ for } j = 2, \ldots, n_i, \\ & \quad (4.4) \end{aligned}$$

where

$$\max\left[\frac{-\mu_{ij}(\cdot)}{1 - \mu_{i,j-1}(\cdot)}, -\frac{1 - \mu_{ij}(\cdot)}{\mu_{i,j-1}(\cdot)}\right] \leq \rho \leq \min\left[\frac{1 - \mu_{ij}(\cdot)}{1 - \mu_{i,j-1}(\cdot)}, \frac{\mu_{ij}(\cdot)}{\mu_{i,j-1}(\cdot)}\right],$$

for $j = 2, \ldots, n_i;\ i = 1, \ldots, K$. In (4.3) and (4.4) $\psi(z_{ij})$ is the non-parametric function added to explain the effect of the secondary covariates $z_{ij}(t_{ij})$ on the binary responses.

Note that the semi-parametric binary correlation model (4.4) is similar but different than the semi-parametric model for the longitudinal count data (Chapter 2). The count data model is based on so-called binary thinning operation (Sutradhar, 2003, McKenzie, 1988). While the binary data model follows a linear correlation model studied earlier by Zeger et al. (1985). In the following subsection, we provide the marginal correlation properties of the SLDCP model (4.3)−(4.4).

4.1.1 Basic properties of the SLDCP model

Lemma 4.1. *Under the SLDCP (4.3)−(4.4), the responses have the following moment properties:*

$$
\begin{aligned}
& \mathrm{E}[Y_{ij}|\boldsymbol{x}_{ij}, z_{ij}] = \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))\ \textit{for all}\ j = 1, \ldots, n_i\,, \\
& \mathrm{Var}[Y_{ij}|\boldsymbol{x}_{ij}, z_{ij}] = \sigma_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) \\
= \ & \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))[1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))],\ \textit{for all}\ j = 1, \ldots, n_i,
\end{aligned} \tag{4.5}
$$

and

$$
\begin{aligned}
& \mathrm{Corr}(Y_{ij}, Y_{ik}|\boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, z_{ij}, z_{ik}) \\
= & \begin{cases} \rho^{k-j}\sqrt{\frac{\mu_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij};\psi(z_{ij}))[1-\mu_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij};\psi(z_{ij}))]}{\mu_{ik}(\boldsymbol{\beta},\boldsymbol{x}_{ik};\psi(z_{ik}))[1-\mu_{ik}(\boldsymbol{\beta},\boldsymbol{x}_{ik};\psi(z_{ik}))]}} & j < k \\ \rho^{j-k}\sqrt{\frac{\mu_{ik}(\boldsymbol{\beta},\boldsymbol{x}_{ik};\psi(z_{ik}))[1-\mu_{ik}(\boldsymbol{\beta},\boldsymbol{x}_{ik};\psi(z_{ik}))]}{\mu_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij};\psi(z_{ij}))[1-\mu_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij};\psi(z_{ij}))]}} & j > k. \end{cases}
\end{aligned} \tag{4.6}
$$

Proof. $\mathrm{E}[Y_{i1}] = \mu_{i1}(\boldsymbol{\beta}, \boldsymbol{x}_{i1}, \psi(z_{i1}))$ follows (4.3). For $j = 2, \ldots, n_i$, by applying (4.4) recursively, the marginal mean property is derived as

$$\begin{aligned}
\mathrm{E}[Y_{ij}] &= \mathrm{E}[\mathrm{E}[Y_{ij}|y_{i,j-1}]] = \mu_{ij} + \rho\,\mathrm{E}[Y_{i,j-1} - \mu_{i,j-1}] \\
&= \mu_{ij} + \rho\,\mathrm{E}[\mathrm{E}[Y_{i,j-1} - \mu_{i,j-1}|y_{i,j-2}]] = \mu_{ij} + \rho^2\,\mathrm{E}[Y_{i,j-2} - \mu_{i,j-2}] \\
&\ \ \vdots \\
&= \mu_{ij} + \rho^{j-1}\,\mathrm{E}[Y_{i1} - \mu_{i1}] \\
&= \mu_{ij}\,.
\end{aligned}$$

The variance follows by definition of the binary response.

For $j < k$ (Sutradhar, 2011, Chapter 7),

$$\begin{aligned}
\mathrm{Cov}[Y_{ij}, Y_{ik}] &= \mathrm{E}[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})] = \mathrm{E}[\mathrm{E}[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})|y_{ij}, \cdots, y_{i,k-1}]] \\
&= \rho\,\mathrm{E}[(Y_{ij} - \mu_{ij})(Y_{i,k-1} - \mu_{i,k-1})] \\
&= \rho\,\mathrm{E}[\mathrm{E}[(Y_{ij} - \mu_{ij})(Y_{i,k-1} - \mu_{i,k-1})|y_{ij}, \cdots, y_{i,k-2}]] \\
&= \rho^2\,\mathrm{E}[(Y_{ij} - \mu_{ij})(Y_{i,k-2} - \mu_{i,k-2})] \\
&\ \ \vdots \quad \text{by applying (4.4) recursively} \\
&= \rho^{k-j}\,\mathrm{E}[(Y_{ij} - \mu_{ij})(Y_{ij} - \mu_{ij})] \\
&= \rho^{k-j}\,\sigma_{ijj}\,,
\end{aligned}$$

which further gives (4.6). □

4.1.2 Estimation for the proposed SLDCP model

Fitting the SLDCP model (4.4) to the repeated binary data requires the estimation of the nonparametric function $\psi(\cdot)$, and the model parameters $\boldsymbol{\beta}$ and ρ. We provide their step by step consistent estimation as follows.

4.1.2.1 SQL estimation of the nonparametric function $\psi(\cdot)$ under the SLDCP model

Note that it is of primary interest to estimate the regression effects $\boldsymbol{\beta}$ involved in the SLDCP model (4.4) consistently and as efficiently as possible. However, a consistent estimator of $\boldsymbol{\beta}$ can not be obtained without consistently estimating the function $\psi(z_{ij})$ involved in the model [see (4.3)]. Thus, for known $\boldsymbol{\beta}$, we first develop a consistent estimator $\hat{\psi}(\boldsymbol{\beta}, z_{ij})$ for $\psi(z_{ij})$. Further, note that given $\boldsymbol{\beta}$, a consistent estimator of $\psi(\cdot)$ can be obtained through the means and variances of the repeated responses $\{y_{ij}, j = 1, \ldots, n_i\}$ only. Afterwards, in a second stage, one can pretend that the repeated responses are independent, that is, assume that $\rho = 0$ in model (4.4). This is equivalent to use the QL (quasi-likelihood) approach (Wedderburn, 1974) which is further equivalent to the independence (I) assumption based GEE (GEE(I)) approach (Liang and Zeger, 1986). In the present context, this QL approach will be referred to as the semi-parametric QL (SQL) approach which we construct as follows.

Without loss of generality, use z_0 for z_{ij} for given i and j, and hence estimate $\psi(z_0)$ at all possible values of z_0 corresponding to all i and j. Note that it is impossible to estimate the regression parameters $\boldsymbol{\beta}$ and the nonparametric function $\psi(z_0)$ independently. Thus, for known $\boldsymbol{\beta}$, under the assumption that the mean function μ_{ij} in (4.5) [see also (4.3)] is continuous, one may solve the SQL (semi-parametric quasi-likelihood) estimating equation for $\psi(z_{ij})|_{z_{ij}=z_0}$ as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\, \frac{\partial \mu_{ij}}{\partial \psi(z_0)} \left(\frac{y_{ij}-\mu_{ij}}{\sigma_{i,jj}}\right) = 0, \tag{4.7}$$

where $\sigma_{i,jj}$ is the variance of y_{ij} as given by (4.5), and

$$w_{ij}(z_0) = p_{ij}(\frac{z_0 - z_{ij}}{b})/\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(\frac{z_0 - z_{ij}}{b}) \tag{4.8}$$

is a kernel weight with p_{ij} as the kernel density, as discussed in (3.12)−(3.14). Note that for $w_{ij}(z_0) = 1$, the SQL estimating equation (4.7) reduces to the well known QL (quasi-likelihood) equation (Wedderburn, 1974). For the bandwidth parameter b in the kernel density $p_{ij}((z_0 - z_{ij})/b)$, we assume that this parameter is chosen such that the mean squared error of the estimator of $\psi(z_{ij})$ will be minimum. By this token, b may be optimally chosen as $b \propto K^{-1/5}$ (Pagan and Ullah, 1999). More specifically, one may use $b = c_0 K^{-1/5}$ where the constant c_0 can be estimated, for example, following Horowitz (2009, Section 2.7) and Powell and Stoker (1996).

Next because

$$\frac{\partial \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))}{\partial \psi(z_0)} = \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))[1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))] = \sigma_{ijj}(z_0),$$

the SQL estimating equation (4.7) reduces to

$$f(\psi(z_0), \boldsymbol{\beta}) = \sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[y_{ij} - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))] = 0, \tag{4.9}$$

which may be solved for the estimate $\hat{\psi}(z_0, \boldsymbol{\beta})$ of $\psi(z_0)$ by using the iterative equation

$$\begin{aligned} \hat{\psi}(z_0, \boldsymbol{\beta})_{(r+1)} &= \hat{\psi}(z_0, \boldsymbol{\beta})_{(r)} \\ &- \left[\{f'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta})\}^{-1} f(\psi(z_0), \boldsymbol{\beta})\right]_{|\psi(z_0) = \hat{\psi}(z_0, \boldsymbol{\beta})_{(r)}}, \end{aligned} \tag{4.10}$$

where (r) indicates the rth iteration, and $f'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta})$ has the formula

$$\begin{aligned} & f'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta}) \\ = & -\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))\{1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))\}]. \end{aligned} \tag{4.11}$$

Note that as shown in Section 4.1.3.1, the SQL estimator $\hat{\psi}(z_0, \boldsymbol{\beta})$ obtained from (4.9) is a consistent estimator for the true nonparametric function $\psi(z_0)$. More specifically, it is shown that $\hat{\psi}(z_0, \boldsymbol{\beta})$ converges to $\psi(z_0)$ provided $Kb^4 \to 0$ as $K \to \infty$, implying that for a constant c^*, the bandwidth parameter b may be chosen as $b = c^* K^{-\alpha}$ with $\alpha > \frac{1}{4}$. A similar result with $\frac{1}{4} < \alpha \leq \frac{1}{3}$ is available in Lin and Carroll (2001), for example. Notice that this choice of b is not optimal, because the asymptotic convergence was derived, for simplicity, only by reducing the bias of the estimator, while the derivation of the optimal choice requires both bias reduction and variance minimization of the estimator.

Note that some authors such as Lin and Carroll (2001) have estimated $\psi(\cdot)$ using 'working' correlations, whereas we have used independence approach to construct the SQL estimating equation (4.7). This is because the SQL estimate from (4.9) is simple and, as shown in Section 4.1.3.1, it is also consistent. It is further seen from Lin and Carroll (2001, Section 4) that the use of such working correlations does not improve the efficiency for the estimates of the main regression parameters. We discuss this issue in details in Section 4.1.4.1.

4.1.2.2 SGQL estimation of the regression effects $\boldsymbol{\beta}$

For known $\boldsymbol{\beta}$, in the last section, we have obtained the SQL estimator $\hat{\psi}(\boldsymbol{\beta}, z_{ij})$ which is a consistent estimator for the true nonparametric function $\psi(z_{ij})$. We now replace the $\psi(z_{ij})$ function involved in the original mean, variance and covariance of the responses given in (4.5)−(4.6) with this estimator $\hat{\psi}(\boldsymbol{\beta}, z_{ij})$, and re-express these moments as

$$\begin{aligned}\tilde{\mu}_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})) &= \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))_{|\psi(z_{ij})=\hat{\psi}(\boldsymbol{\beta}, z_{ij})} \\ \tilde{\sigma}_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})) &= \sigma_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))_{|\psi(z_{ij})=\hat{\psi}(\boldsymbol{\beta}, z_{ij})},\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}&\tilde{\sigma}_{i,jk}(\boldsymbol{\beta}, \rho, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}), \hat{\psi}(\boldsymbol{\beta}, z_{ik})) \\ &= \sigma_{i,jk}(\boldsymbol{\beta}, \rho, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \psi(z_{ij}), \psi(z_{ik}))_{|\psi(z_{ij})=\hat{\psi}(\boldsymbol{\beta}, z_{ij}), \psi(z_{ik})=\hat{\psi}(\boldsymbol{\beta}, z_{ik})} \text{ for } j < k,\end{aligned} \tag{4.13}$$

respectively. Further, for notational simplicity, in the rest of this section, we use $\tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))$ for $\tilde{\mu}_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))$; $\tilde{\sigma}_{i,jj}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))$ for $\tilde{\sigma}_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))$; and $\tilde{\sigma}_{i,jk}(\boldsymbol{\beta}, \rho, \hat{\psi}(\boldsymbol{\beta}))$ for $\tilde{\sigma}_{i,jk}(\boldsymbol{\beta}, \rho, \boldsymbol{x}_{ij}, \boldsymbol{x}_{ik}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}), \hat{\psi}(\boldsymbol{\beta}, z_{ik}))$. We now express the mean and covariance matrix of the binary response vector $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})^\top$ as

$$\begin{aligned}&\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta})) = E[\boldsymbol{Y}_i] = [\tilde{\mu}_{i1}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta})), \ldots, \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta})), \ldots, \tilde{\mu}_{in_i}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))]^\top : n_i \times 1 \\ &\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta}, \rho, \hat{\psi}(\boldsymbol{\beta})) = \text{Cov}[\boldsymbol{Y}_i] = (\tilde{\sigma}_{i,jk}(\boldsymbol{\beta}, \rho, \hat{\psi}(\boldsymbol{\beta}))) : n_i \times n_i.\end{aligned} \tag{4.14}$$

Next, following Sutradhar (2003, 2010), for example, we construct the GQL (generalized quasi-likelihood) estimating equation for $\boldsymbol{\beta}$ as

$$\sum_{i=1}^{K} \frac{\partial [\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))]^\top}{\partial \boldsymbol{\beta}} \; [\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta}, \rho, \hat{\psi}(\boldsymbol{\beta}))]^{-1} \; [\boldsymbol{y}_i - \tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))] = 0, \tag{4.15}$$

where

$$\frac{\partial [\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))]^\top}{\partial \boldsymbol{\beta}} = \frac{\partial (\tilde{\mu}_{i1}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta})), \ldots, \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta})), \ldots, \tilde{\mu}_{in_i}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta})))}{\partial \boldsymbol{\beta}}.$$

We remark that to reflect the semi-parametric means and correlations involved in the GQL estimating equation (4.15) for $\boldsymbol{\beta}$, we refer to this estimating equation as the semi-parametric GQL (SGQL) estimating equation. Let the solution of (4.15) be denoted by $\hat{\boldsymbol{\beta}}_{SGQL}$. Further note that in the derivative matrix in the SGQL estimating equation (4.15), the derivative vector for the jth element has the formula

$$\begin{aligned}
&\frac{\partial \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} = \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))(1 - \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))) \left(\boldsymbol{x}_{ij} + \frac{\partial}{\partial \boldsymbol{\beta}} \hat{\psi}(\boldsymbol{\beta}, z_{ij})\right) \\
= \;& \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))(1 - \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))) \left[\boldsymbol{x}_{ij} \right. \qquad (4.16) \\
- \;& \left. \frac{\sum_{\ell=1}^{K} \sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij}) \tilde{\mu}_{\ell u}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))(1 - \tilde{\mu}_{\ell u}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))) \boldsymbol{x}_{\ell u}}{\sum_{\ell=1}^{K} \sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij}) \tilde{\mu}_{\ell u}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))(1 - \tilde{\mu}_{\ell u}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}, z_{ij})))} \right],
\end{aligned}$$

where $\hat{\psi}(\boldsymbol{\beta}, z_{ij})$ is the solution of (4.9).

Thus, the $\hat{\boldsymbol{\beta}}_{SGQL}$ estimate may be obtained by solving (4.15) iteratively until convergence. As we will see later, this estimator is consistent for true $\boldsymbol{\beta}$ and it is always more efficient than the existing GEE(I), i.e. SQL approaches, whereas the so-called GEE approaches may not satisfy this fundamental inequality. The asymptotic convergence is explained in Section 4.1.3.2 and its finite sample performance both for bias and efficiency is discussed through a simulation study in Section 4.1.4.2.

Note that for efficient estimation of the regression effects (of the primary covariates), some authors (Severini and Staniswalis, 1994, Lin and Carroll, 2001) used the so-called unstructured (UNS) correlation matrix based GEE approach. As an extension of the longitudinal model based study (Sutradhar and Das, 1999) to the semi-parametric longitudinal setup, we show in Section 4.1.4.1 through an empirical study that the semi-parametric GEE(UNS) (SGEE (UNS)) approach used in Lin and Carroll (2001) may produce inefficient regression estimates, as compared to the semi-parametric QL (SQL) or SGEE(I) approach. Thus, the SGEE approach can not be recommended in practice for regression estimation.

4.1.2.3 Semi-parametric method of moments (SMM) estimation for the correlation index parameter ρ

When the regression effects $\boldsymbol{\beta}$ and the nonparametric function $\psi(z_{ij})$ are known, one may use the method of moments and exploit the second order moments from (4.5) and (4.6) and obtain the moment estimator of ρ as

$$\hat{\rho} \;=\; \frac{\sum_{i=1}^{K}\sum_{j=2}^{n_i} y^*_{ij}y^*_{i,j-1}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} y^{*2}_{ij}} \; \frac{\sum_{i=1}^{K} n_i}{\sum_{i=1}^{K}\sum_{j=2}^{n_i}\left[\frac{\sigma_{i,j-1,j-1}}{\sigma_{i,jj}}\right]^{\frac{1}{2}}} \tag{4.17}$$

(Sutradhar and Kovacevic, 2000; Sutradhar, 2011, Eqn. (7.88)], where

$$\begin{aligned} y^*_{ij} &= [y_{ij} - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))]/\sqrt{\sigma_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))} \text{ with} \\ \sigma_{i,jj}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) &= \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))(1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))), \end{aligned}$$

and $\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) = \exp\{\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \psi(z_{ij})\}/[1 + \exp\{\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \psi(z_{ij})\}]$.

Next by replacing $\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))$ with

$$\bar{\bar{\mu}}_{ij}(\hat{\boldsymbol{\beta}}_{SGQL}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij})) = \frac{\exp\{\boldsymbol{x}_{ij}^{\top}\hat{\boldsymbol{\beta}}_{SGQL} + \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij})\}}{1 + \exp\{\boldsymbol{x}_{ij}^{\top}\hat{\boldsymbol{\beta}}_{SGQL} + \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij})\}}$$

in (4.17), we obtain the SMM (semi-parametric method of moment) estimator of ρ as

$$\hat{\bar{\rho}} \;=\; \frac{\sum_{i=1}^{K}\sum_{j=2}^{n_i} \bar{y}^*_{ij}\bar{y}^*_{i,j-1}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} \bar{y}^{*2}_{ij}} \; \frac{\sum_{i=1}^{K} n_i}{\sum_{i=1}^{K}\sum_{j=2}^{n_i}\left[\frac{\bar{\bar{\sigma}}_{i,j-1,j-1}}{\bar{\bar{\sigma}}_{i,jj}}\right]^{\frac{1}{2}}}, \tag{4.18}$$

where

$$\begin{aligned} \bar{y}^*_{ij} &= \frac{y_{ij} - \bar{\bar{\mu}}_{ij}(\hat{\boldsymbol{\beta}}_{SGQL}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij}))}{\sqrt{\bar{\bar{\sigma}}_{i,jj}(\hat{\boldsymbol{\beta}}_{SGQL}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij}))}} \\ \bar{\bar{\sigma}}_{i,jj}(\hat{\boldsymbol{\beta}}_{SGQL}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij})) &= \bar{\bar{\mu}}_{ij}(\hat{\boldsymbol{\beta}}_{SGQL}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij})) \end{aligned}$$

$$\times \quad [1 - \bar{\bar{\mu}}_{ij}(\hat{\boldsymbol{\beta}}_{SGQL}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}_{SGQL}, z_{ij}))].$$

The consistency of this SMM estimator $\hat{\hat{\rho}}$ will be shown in brief in Section 4.1.3.3.

4.1.3 Asymptotic properties of the estimators of the SLDCP model

4.1.3.1 Consistency of $\hat{\psi}(\cdot)$

For convenience, in (4.9), we have shown the estimation for $\psi(z_0)$ for $z_0 = z_{\ell u}$ for a selected value of $\ell(\ell = 1, \ldots, K)$ and $u(u = 1, \ldots, n_\ell)$. For notational simplicity, here we use $\mu_{ij}(z_0)$ for $\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_0))$. Now, for known $\boldsymbol{\beta}$, and for true binary mean $\mu_{ij} \equiv \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij}))$ given by (4.3), with the similar idea as in Section 3.3.1, a Taylor expansion of $f(\hat{\psi}(z_0; \boldsymbol{\beta}), \boldsymbol{\beta}) = \sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z_0)[y_{ij} - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(z_0; \boldsymbol{\beta}))]$ (4.9) about $\psi(z_0)$ gives

$$\begin{aligned}
\hat{\psi}(z_0; \boldsymbol{\beta}) - \psi(z_0) &\approx \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z_0)\left[y_{ij} - \mu_{ij}(z_0)\right]}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z_0)\mu_{ij}(z_0)\left[1 - \mu_{ij}(z_0)\right]} = \frac{f(\psi(z_0), \boldsymbol{\beta})}{f'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta})} \\
&= \frac{1}{f'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta})}\left[\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z_0)\left[y_{ij} - \mu_{ij}\right] + \sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z_0)\left[\mu_{ij} - \mu_{ij}(z_0)\right]\right] \\
&= A_K + H_K\,, \qquad (4.19)
\end{aligned}$$

where

$$\begin{aligned}
H_K &= \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\left[\mu_{ij} - \mu_{ij}(z_0)\right], \text{ and} \\
A_K &= \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\left(y_{ij} - \mu_{ij}\right) \text{ with} \\
B_K &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\mu_{ij}(z_0)\left[1 - \mu_{ij}(z_0)\right],
\end{aligned}$$

and $p_{ij}(z_0) \equiv p_{ij}(\frac{z_0 - z_{ij}}{b})$ being the kernel density defined in (3.13). Here $b \propto K^{-\alpha}$ for a suitable α (Pagan and Ullah, 1999, p. 28; Lin and Carroll, 2001). The asymptotic behaviors of A_K and H_K are given by the following lemmas.

Lemma 4.2.

$$A_K = O_p(1/\sqrt{K}). \tag{4.20}$$

Proof. Notice that $\mathrm{E}[A_K] = 0$ and

$$\mathrm{Var}[A_K] = \frac{1}{B_K^2}\frac{1}{K^2}\sum_{i=1}^{K} \mathrm{Var}\left[\sum_{j=1}^{n_i} p_{ij}(z_0)Y_{ij}\right] = \frac{1}{K}Q_k$$

with $Q_K = \frac{1}{B_K^2}\frac{1}{K}\sum_{i=1}^{K}\mathrm{Var}\left[\sum_{j=1}^{n_i} p_{ij}(z_0)Y_{ij}\right] = O(1)$. The result in (4.20) then follows, for example, from Amemiya (1985, Theorem 14.4-1). □

Lemma 4.3.

$$H_K = O(b^2), \tag{4.21}$$

where b is the bandwidth parameter involved in the kernel density.

Proof. By using

$$\mu_{ij} - \mu_{ij}(z_0) = \mu_{ij}(z_0)\left[1 - \mu_{ij}(z_0)\right]\psi'(z_0)(z_{ij} - z_0) + O\left((z_{ij} - z_0)^2\right),$$

we write

$$H_K \approx \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\,\mu_{ij}(z_0)\left[1 - \mu_{ij}(z_0)\right]\psi'(z_0)\,(z_{ij} - z_0)$$

$$
\begin{aligned}
&= \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\mu_{ij}(z_0)\left[1-\mu_{ij}(z_0)\right]\psi'(z_0)\left\{p_{ij}(z_0)\left(z_{ij}-z_0\right)-\mathrm{E}\left[p_{ij}(z_0)\left(z_{ij}-z_0\right)|\mathbf{x}_{ij}\right]\right\} \\
&\quad + \frac{1}{B_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\mu_{ij}(z_0)\left[1-\mu_{ij}(z_0)\right]\psi'(z_0)\,\mathrm{E}\left[p_{ij}(z_0)\left(z_{ij}-z_0\right)|\mathbf{x}_{ij}\right] \\
&= O(b^2),
\end{aligned}
$$

according to result (3.58), which is applicable to this SLDCP setup. For the first term, due to $p_{ij}(z_0)$, its variance is in the order of $O(b^2/K)$, so it is $O_p(b/\sqrt{K})$, which can be neglected. Thus we have shown that $H_K = O(b^2)$. □

By using (4.20) and (4.21) in (4.19), one obtains

$$
\hat{\psi}(z_0;\boldsymbol{\beta}) - \psi(z_0) = A_K + O(b^2) = O_p(1/\sqrt{K}) + O(b^2), \tag{4.22}
$$

showing that $\hat{\psi}(z_0;\boldsymbol{\beta})$ is consistent for $\psi(z_0)$ provided $Kb^4 \to 0$ as $K \to \infty$, that is, $K\frac{1}{K^{4\alpha}} = \frac{1}{K^{4\alpha-1}} \to 0$, yielding the condition $\alpha > 1/4$. Note that this convergence result is obtained by minimizing the bias of the estimator [see (4.19)] only.

4.1.3.2 Asymptotic normality and consistency of $\hat{\boldsymbol{\beta}}_{SGQL}$

The asymptotic result of the SGQL estimator $\hat{\boldsymbol{\beta}}_{SGQL}$ of $\boldsymbol{\beta}$ is given by the following lemma.

Lemma 4.4.

$$
\begin{aligned}
\sqrt{K}\left\{\hat{\boldsymbol{\beta}}_{SGQL} - \boldsymbol{\beta}\right\} &= \boldsymbol{F}^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i}\right)\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right) \\
&+ O(\sqrt{Kb^4}) + o_p(1),
\end{aligned} \tag{4.23}
$$

where $\boldsymbol{Z}_{1i} = \frac{\partial \tilde{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} \tilde{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}), \rho)$, *and* $\boldsymbol{Z}_{2i} = (\boldsymbol{Z}_{2i1}, \cdots, \boldsymbol{Z}_{2in_i})$ *with*

$$\begin{aligned} \boldsymbol{Z}_{2ij} &= \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_i}\sum_{k'=1}^{n_i} \frac{1}{B_K(z_{i'k'})} \frac{\partial \tilde{\mu}_{i'j'}(\boldsymbol{\beta}, \hat{\psi}(z_{i'j'};\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} v_{1i'}^{j'k'}(\boldsymbol{\beta}, \hat{\psi}, \rho) \\ &\times \ \mu_{i'k'}(\boldsymbol{\beta}, \psi(z_{i'k'}))\left[1 - \mu_{i'k'}(\boldsymbol{\beta}, \psi(z_{i'k'}))\right] p_{ij}(z_{i'k'}), \end{aligned}$$

where B_K *and the kernel density* $p_{ij}(z_0)$ *are defined in (4.19), and* v_{1i}^{jk} *is the* (j,k)*th element of* $\tilde{\boldsymbol{\Sigma}}_i^{-1}$*.*

Proof. Recall that the SGQL estimator $\hat{\boldsymbol{\beta}}_{SGQL}$ of $\boldsymbol{\beta}$ is obtained by solving the estimating equation (4.15). For true $\boldsymbol{\beta}$, denote the estimating function in (4.15) as

$$\boldsymbol{D}_K(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K} \frac{\partial[\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))]^\top}{\partial \boldsymbol{\beta}} \ [\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta}, \rho, \hat{\psi}(\boldsymbol{\beta}))]^{-1} \ [\boldsymbol{y}_i - \tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))].$$

Thus, $\hat{\boldsymbol{\beta}}_{SGQL}$ must satisfy $\boldsymbol{D}_K(\hat{\boldsymbol{\beta}}_{SGQL}) = 0$, which by a linear Taylor expansion about true $\boldsymbol{\beta}$ provides

$$\boldsymbol{D}_K(\boldsymbol{\beta}) + (\hat{\boldsymbol{\beta}}_{SGQL} - \boldsymbol{\beta})\boldsymbol{D}'_K(\boldsymbol{\beta}) + o_p(1/\sqrt{K}) = 0. \tag{4.24}$$

Thus,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{SGQL} - \boldsymbol{\beta} &= -\left[\boldsymbol{D}'_K(\boldsymbol{\beta})\right]^{-1}\left[\boldsymbol{D}_K(\boldsymbol{\beta}) + o_p(1/\sqrt{K})\right] \\ &= \left[\boldsymbol{F}_K(\boldsymbol{\beta})\right]^{-1}\boldsymbol{D}_K(\boldsymbol{\beta}) + o_p(1/\sqrt{K}), \end{aligned} \tag{4.25}$$

where

$$\boldsymbol{F}_K(\boldsymbol{\beta}) = \frac{1}{K}\sum_{i=1}^{K} \frac{\partial[\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))]^\top}{\partial \boldsymbol{\beta}} \ [\tilde{\boldsymbol{\Sigma}}_i(\boldsymbol{\beta}, \rho, \hat{\psi}(\boldsymbol{\beta}))]^{-1} \ \frac{\partial \tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^\top}.$$

Notice that in (4.25), one may write

$$\boldsymbol{F}(\boldsymbol{\beta}) = \lim_{K\to\infty} \boldsymbol{F}_K(\boldsymbol{\beta}) = \mathrm{E}_{\hat{\psi}} \left[\frac{\partial \tilde{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} \tilde{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}), \rho) \frac{\partial \tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^\top} \right].$$

Next the estimating function $\boldsymbol{D}_K(\boldsymbol{\beta})$ in (4.25) may be further expressed as

$$\begin{aligned}
\boldsymbol{D}_K(\boldsymbol{\beta}) &= \frac{1}{K}\sum_{i=1}^K \boldsymbol{Z}_{1i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) \\
&- \frac{1}{K}\sum_{i=1}^K \frac{\partial \tilde{\boldsymbol{\mu}}_i^\top(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} \tilde{\boldsymbol{\Sigma}}_i^{-1}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta}), \rho) \left[\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{z}_i;\boldsymbol{\beta})) - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \psi(\boldsymbol{z}_i))\right] \\
&= \frac{1}{K}\sum_{i=1}^K \boldsymbol{Z}_{1i}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) - \frac{1}{K}\sum_{i=1}^K\sum_{j=1}^{n_i}\sum_{k=1}^{n_i} \frac{\partial \tilde{\mu}_{ij}(\boldsymbol{\beta}, \hat{\psi}(z_{ij};\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}} v_{1i}^{jk}(\boldsymbol{\beta}, \hat{\psi}, \rho) \\
&\times \mu_{ik}(\boldsymbol{\beta}, \psi(z_{ik}))\left[1 - \mu_{ik}(\boldsymbol{\beta}, \psi(z_{ik}))\right]\left[\hat{\psi}(z_{ik};\boldsymbol{\beta}) - \psi(z_{ik})\right] + o_p(1/\sqrt{K}) \\
&= \frac{1}{K}\sum_{i=1}^K (\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i})(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) + O(b^2) + o_p(1/\sqrt{K}),
\end{aligned} \tag{4.26}$$

by (4.22).

Hence by using (4.26) in (4.25), one obtains the result (4.23). □

Theorem 4.1. *The SGQL estimator* $\hat{\boldsymbol{\beta}}_{SGQL}$ *(the solution of (4.15)) has the limiting (as* $K \to \infty$*) multivariate normal distribution given as*

$$\sqrt{K}\left\{\hat{\boldsymbol{\beta}}_{SGQL} - \boldsymbol{\beta} - O(b^2)\right\} \;\to\; N(0, \boldsymbol{V}_{\boldsymbol{\beta}}), \tag{4.27}$$

where

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{F}^{-1}\frac{1}{K}\left[\sum_{i=1}^K (\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i})\boldsymbol{\Sigma}_i(\boldsymbol{Z}_{1i} - \boldsymbol{Z}_{2i})^\top\right]\boldsymbol{F}^{-1}.$$

Proof. Because $E[\boldsymbol{Y}_i - \boldsymbol{\mu}_i] = 0$, and $\mathrm{cov}[\boldsymbol{Y}_i] = \boldsymbol{\Sigma}_i$, under the conditions (3.76), by applying Lindeberg-Feller central limit theorem (Bishop, Fienberg and Holland, 2007, Theorem 3.3.6) for independent random variables with non-identical distributions to

Lemma 4.4, one obtains the theorem. □

4.1.3.3 Consistency of $\hat{\hat{\rho}}$

We proved the consistency for known $\boldsymbol{\beta}$ and $\psi(z_{ij})$. The result remains valid when $\boldsymbol{\beta}$ and $\psi(z_{ij})$ are replaced by their respective consistent estimates. The consistency of the moment estimator $\hat{\rho}$ (4.17) is given by the following lemma:

Lemma 4.5. *The moment estimator $\hat{\rho}$ given in (4.17) is a consistent estimator for the longitudinal correlation index parameter ρ.*

Proof. Notice that for known $\boldsymbol{\beta}$ and $\psi(z_{ij})$, the moment estimator of ρ is given by (4.17). For two fixed quantities M_1 and M_2, we assume that the lag 1 sum of products and sum of squares used in (4.17) have bounded variances satisfying

$$\mathrm{E}\left[\left(\sum_{j=1}^{n_i-1}\left[Y_{ij}^{*}Y_{i,j+1}^{*}-\frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right]\right)^2\right] \;<\; M_1, \text{ and}$$

$$\mathrm{E}\left[\left(\sum_{j=1}^{n_i}\left[Y_{ij}^{*2}-1\right]\right)^2\right] \;<\; M_2,$$

respectively. Now because Y_{ij}'s are independent for different i, for $K \to \infty$, we may apply the law of large numbers for independent random variables (Breiman, 1968, Theorem 3.27) and obtain

$$\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\left(y_{ij}^{*}y_{i,j+1}^{*}-\frac{\sigma_{i,j,j+1}}{\sqrt{\sigma_{ijj}\sigma_{i,j+1,j+1}}}\right)}{\sum_{i=1}^{K}(n_i-1)} \xrightarrow{P} 0$$

$$\Rightarrow \quad \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1} y_{ij}^{*}y_{i,j+1}^{*}}{\sum_{i=1}^{K}(n_i-1)} = \frac{\rho\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\frac{\sqrt{\sigma_{ijj}}}{\sqrt{\sigma_{i,j+1,j+1}}}}{\sum_{i=1}^{K}(n_i-1)} + o_p(1). \qquad (4.28)$$

Similarly,

$$\begin{aligned}
\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(y_{ij}^{*\,2}-1\right)}{\sum_{i=1}^{K} n_i} &\xrightarrow{P} 0 \qquad \Rightarrow \\
\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}^{*\,2}}{\sum_{i=1}^{K} n_i} &= 1+o_p(1). \qquad (4.29)
\end{aligned}$$

Dividing (4.28) by (4.29), after some algebra, one obtains

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1} y_{ij}^{*} y_{i,j+1}^{*}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}^{*\,2}} \frac{\sum_{i=1}^{K} n_i}{\sum_{i=1}^{K}\sum_{j=1}^{n_i-1}\left[\frac{\sigma_{ijj}}{\sigma_{i,j+1,j+1}}\right]^{\frac{1}{2}}} \xrightarrow{P} \rho \quad \text{as } K \to \infty\,. \qquad (4.30)$$

$\square$

The consistency for $\hat{\hat{\rho}}$ in (4.18) follows from (4.30) because of the fact that $\hat{\hat{\rho}}$ was constructed by putting consistent estimates for $\boldsymbol{\beta}$ and $\psi(z_{ij})$ in the formula for $\hat{\rho}$ in (4.17).

4.1.4 A simulation study

The main objective of this simulation study (see Section 4.1.4.2) is to examine the finite sample bias and efficiency performance of the proposed SGQL estimator of the regression parameter $\boldsymbol{\beta}$ obtained by solving the SGQL estimating equation (4.15). Because this $\boldsymbol{\beta}$ parameter is involved in the AR(1) (auto-regressive order 1) type SLDCP model (4.4) for repeated binary data, it can not be estimated without estimating the nonparametric function $\psi(\cdot)$ and the longitudinal correlation index parameter ρ. The estimates in the simulation study are obtained by solving the SQL estimating equation (4.9) for the $\psi(\cdot)$ function and the SMM equation (4.18) for the ρ parameter. Note that only the SGQL estimates of the main parameter $\boldsymbol{\beta}$ are compared with the existing GEE estimates obtained by using 'working' MA(1) (moving average of order 1), EQC (equi-correlations) and independence (I) assumption under a truly AR(1) binary data. The unstructured (UNS) 'working' correlation structure (Severini and Staniswalis, 1994, Lin and Carroll, 2001) is not used in this simulation study in Section 4.1.4.2, because, as we demonstrate in Section 4.1.4.1, the GEE(UNS) approach may produce less efficient estimates than the GEE(I) (independence assumption based GEE) estimates which makes the GEE approach useless. Nevertheless, other possible 'working' correlations (MA(1), EQC, I) based GEE were included in Section 4.1.4.2 for the sake of completeness.

As far as the primary and secondary covariates are concerned, we choose the primary covariates as:

For $K = 50$,

$$
\begin{aligned}
x_{ij1}(j) &= \begin{cases} \frac{1}{2} & \text{for } i = 1, \ldots, 10 \text{ and } j = 1, 2 \\ 1 & \text{for } i = 1, \ldots, 10 \text{ and } j = 3, 4 \\ -\frac{1}{2} & \text{for } i = 11, \ldots, 40 \text{ and } j = 1 \\ 0 & \text{for } i = 11, \ldots, 40 \text{ and } j = 2, 3 \\ \frac{1}{2} & \text{for } i = 41, \ldots, 50 \text{ and } j = 4 \\ \frac{j}{2n_i} & \text{for } i = 41, \ldots, 50 \text{ and } j = 1, 2, 3, 4 \end{cases} \qquad (4.31) \\
x_{ij2}(j) &= \begin{cases} \frac{j-2.5}{2n_i} & \text{for } i = 1, \ldots, 25 \text{ and } j = 1, 2, 3, 4 \\ 0 & \text{for } i = 26, \ldots, 50 \text{ and } j = 1, 2 \\ \frac{1}{2} & \text{for } i = 26, \ldots, 50 \text{ and } j = 3, 4. \end{cases}
\end{aligned}
$$

For $K = 70$,

$$x_{ij1}(j) = \begin{cases} \frac{1}{2} & \text{for } i = 1, \ldots, 15 \text{ and } j = 1, 2 \\ 1 & \text{for } i = 1, \ldots, 15 \text{ and } j = 3, 4 \\ -\frac{1}{2} & \text{for } i = 16, \ldots, 55 \text{ and } j = 1 \\ 0 & \text{for } i = 16, \ldots, 55 \text{ and } j = 2, 3 \\ \frac{1}{2} & \text{for } i = 16, \ldots, 55 \text{ and } j = 4 \\ \frac{j}{2n_i} & \text{for } i = 56, \ldots, 70 \text{ and } j = 1, 2, 3, 4 \end{cases} \tag{4.32}$$

$$x_{ij2}(j) = \begin{cases} \frac{j-2.5}{2n_i} & \text{for } i = 1, \ldots, 35 \text{ and } j = 1, 2, 3, 4 \\ 0 & \text{for } i = 36, \ldots, 70 \text{ and } j = 1, 2 \\ \frac{1}{2} & \text{for } i = 36, \ldots, 70 \text{ and } j = 3, 4. \end{cases}$$

For $K = 100$: This design is the same as in (3.105).

The secondary covariates (z_{ij}) and nonparametric functions ($\psi(z_{ij})$) are chosen as in (3.106) and (3.107), respectively.

Furthermore, for the bandwidth parameter involved in the kernel weights we choose the recommended optimal value under the independent setup, namely $b = c_0 K^{-1/5}$, where $K = 50, 70,$ or 100. As far as c_0 is concerned, the formula in Horowitz (2009, Section 2.7) appears to be complex. Because this parameter is set for all possible small partitions for the secondary covariate z, we have treated c_0 as the standard deviation of z values from the entire space. For example, in Chapter 3, for z values ranging from 0.5 to 4.5, c_0 was chosen as $c_0 = \sigma_z \approx \text{range}/4 = [4.5 - 0.5]/4 = 1.0$ (see Figure 3.1). This choice of normalizing constant works better than other choices, which we verified by searching for mini-max MSE (mean squared error) of the estimators. We do not report the detailed results here to save space.

4.1.4.1 SGEE estimation of regression parameter $\boldsymbol{\beta}$ and drawbacks

Notice that under the SLDCP model (4.4), $\psi(z_0)$ is estimated by solving (4.9) as a function of $\boldsymbol{\beta}$ by treating $\rho = 0$, ρ being the dynamic dependence parameter of the model. The estimator is denoted by $\hat{\psi}(z_0, \boldsymbol{\beta})$. Next, we remark that for the estimation of the main regression parameter $\boldsymbol{\beta}$, some of the existing studies (Severini and Staniswalis, 1994, Lin and Carroll, 2001) have dealt with marginal models where the mean and the variances are not affected by the dynamic dependence parameter. To accommodate possible correlation of the repeated data, these authors have used a 'working' correlation structure based GEE (generalized estimating equation) approach for efficient estimation of $\boldsymbol{\beta}$ which does not require any modeling for the true correlation structure such as using (4.4). More specifically, using $\boldsymbol{y}_i = [y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}]^\top$, the vector of repeated responses, and $\boldsymbol{z}_i = [z_{i1}, \ldots, z_{ij}, \ldots, z_{in_i}]^\top$, corresponding vector of secondary covariates, the 'working' correlations approach solves the GEE defined as

$$\sum_{i=1}^{K} \frac{\partial \boldsymbol{\mu}_i^\top(\boldsymbol{\beta}, \boldsymbol{X}_i, \hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_i))}{\partial \boldsymbol{\beta}} V_i^{-1}(\boldsymbol{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{X}_i, \hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_i))) = 0, \tag{4.33}$$

where $X_i^\top = (\boldsymbol{x}_i(t_{i1}), \ldots, \boldsymbol{x}_i(t_{ij}), \ldots, \boldsymbol{x}_i(t_{in_i}))$ denote the $p \times n_i$ covariate matrix with $\boldsymbol{x}_i(t_{ij})$ as the $p-$dimensional primary covariate vector as in (4.3) for the ith individual at time point t_{ij}, $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \boldsymbol{X}_i, \hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_i)$ is a mean vector constructed from (4.14)$-$(4.15), where for known $\boldsymbol{\beta}$, $\hat{\psi}(\boldsymbol{\beta}, \boldsymbol{z}_i)$ is a $n_i \times 1$ consistent estimate of the nonparametric vector function $\psi(\boldsymbol{z}_i)$. As opposed to (4.15), the $\boldsymbol{V}_i$ in (4.33) is a so-called $n_i \times n_i$ 'working' correlation matrix representing the correlations of the repeated responses which is computed by $\boldsymbol{V}_i = \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{R}_i \boldsymbol{A}_i^{\frac{1}{2}}$ where $\boldsymbol{A}_i = \text{diag}[\text{var}(y_{i1}), \ldots, \text{var}(y_{ij}), \ldots, \text{var}(y_{in_i})]$ with

$$\begin{aligned}
\text{var}(y_{ij}) &= \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))[1 - \mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))] \\
&= v_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \hat{\psi}(\boldsymbol{\beta}, z_{ij}))
\end{aligned}$$

as in (4.14), but $\boldsymbol{R}_i$ has been computed by an unstructured (UNS) common constant correlation matrix ($\boldsymbol{R}$) as

$$\boldsymbol{R}(\equiv \boldsymbol{R}_i) = K^{-1}\sum_{i=1}^{K}\boldsymbol{r}_i\boldsymbol{r}_i^\top, \text{ where } \boldsymbol{r}_i = (r_{i1},\ldots,r_{ij},\ldots,r_{in_i})^\top, \tag{4.34}$$

with $r_{ij} = \frac{(y_{ij}-\mu_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij},\hat{\psi}(\boldsymbol{\beta},z_{ij})))}{[v_{ij}(\boldsymbol{\beta},\boldsymbol{x}_{ij},\hat{\psi}(\boldsymbol{\beta},z_{ij}))]^{\frac{1}{2}}}$.

Because of the serious inefficiency drawbacks of the GEE approach in the longitudinal setup (Sutradhar and Das, 1999) where GEE was found to be less efficient than using independence approach, it is first worth checking the performance of the existing UNS matrix $\boldsymbol{R}$ based GEE (4.33) (GEE(UNS)) approach (Lin and Carroll, 2001) for estimation of $\boldsymbol{\beta}$ in the present semi-parametric longitudinal setup, before this approach is included in overall comparison under Section 4.1.4.2.

For the purpose we consider $K = 50, 70$, independent individuals each providing repeated binary responses for $n_i = n = 4$ times. We take these individuals as having the two primary covariates with their effects $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (0.5, 0.5)^\top$ on the responses, where the values of the covariates are given as in (4.31) and (4.32). Also the nonparametric function in secondary covariates is given by (3.107). Then, we generate the repeated binary responses $\{y_{ij}, j = 1, \ldots, n_i = n = 4\}$ following the SLDCP model (4.4) using these parameters, covariates and nonparametric function. As far as the correlation index parameter is concerned, we choose $\rho = 0.1$, and the actual correlations among the data were computed by (4.6). However, to estimate $\boldsymbol{\beta}$ by using the UNS 'working' correlation matrix $\boldsymbol{R}$ (4.34) based SGEE in (4.33), one does not need to know this correlation structure (4.6). Following Lin and Carroll (2001, Eqn. (10)), in a given simulation, we obtain the estimate of $\boldsymbol{\beta}$ by solving (4.33), where the function $\psi(\cdot)$ is estimated by (4.9). We use 1000 simulations and denote the average (simulated mean (SM)) of these $\boldsymbol{\beta}$ estimates by $\hat{\boldsymbol{\beta}}_{SGEE(UNS)}$ and also compute the simulated standard error (SSE). In order to examine the relative efficiency performance of this estimate $\hat{\boldsymbol{\beta}}_{SGEE(UNS)}$ with the estimate obtained under the independence assumption ($\boldsymbol{R} = \boldsymbol{I}_4$), we obtain the SGEE(I) estimate by (4.33) but by treating $\boldsymbol{R} = \boldsymbol{I}_4$, which is denoted by $\hat{\boldsymbol{\beta}}_{SGEE(I)}$, and is also the SQL (semi-parametric QL) or moment estimate. As an illustration, we now display these estimates along their simulated mean squared error (SMSE) as follows for $K = 50$ and $K = 70$:

	K=50		K=70	
Quantity	$\hat{\beta}_{1,SGEE(UNS)}$	$\hat{\beta}_{1,SGEE(I)}$	$\hat{\beta}_{1,SGEE(UNS)}$	$\hat{\beta}_{1,SGEE(I)}$
SM	0.574	0.552	0.525	0.509
SSE	0.561	0.533	0.456	0.443
SMSE	0.320	0.286	0.208	0.196
	$\hat{\beta}_{2,SGEE(UNS)}$	$\hat{\beta}_{2,SGEE(I)}$	$\hat{\beta}_{2,SGEE(UNS)}$	$\hat{\beta}_{2,SGEE(I)}$
SM	0.577	0.541	0.568	0.562
SSE	1.207	1.128	0.968	0.922
SMSE	1.461	1.272	0.940	0.852

Table 4.1: Illustration of relative efficiency performance of the SGEE(UNS) (Lin and Carroll (2001)) and SGEE(I) approaches in estimating regression effects $\boldsymbol{\beta}$

Notice that SGEE(I) approach estimates of both β_1 and β_2, have smaller SSE and MSE than the ones obtained via an unstructured (UNS) 'working' correlations based SGEE(UNS) (Lin and Carroll, 2001) approach. For example, when $K = 70$, the SGEE(UNS) produces β_2 estimate with MSE 0.940, while the simpler SGEE(I)≡SQL approach shows the MSE as 0.852. This example, therefore, illustrates that it is useless to attempt applying the 'working' correlations based GEE approach to increase the efficiency in regression estimation because the independence assumption based approach may produce, at times, better estimates. This recommendation for the semi-parametric longitudinal models is similar to that of (Sutradhar and Das, 1999) for the longitudinal models. Nevertheless, for the sake of completeness, we use other 'working' correlations based GEE approaches in the next section to compare their efficiency performance with the proposed SGQL approach.

4.1.4.2 Performance of the proposed SGQL estimation approach

Because in the semi-parametric longitudinal setup, the existing works (Severini and Staniswalis, 1994, Lin and Carroll, 2001, for example) recommended the use of the GEE(UNS) for inferences about the main regression parameters of the model, in the last section we conducted a separate simulation study to examine the performance of this recommended approach. The simulation study however produced contradictions and suggests not to use such GEE(UNS) approach as it fails to gain efficiency at times over the GEE(I) (independence assumption based) approach. In this section, we examine the performance of the proposed SGQL approach in estimating the nonparametric function and parameters of the SLDCP model (4.4). For the sake of completeness we also include some other possible 'working' correlations (other than UNS) based GEE approaches. The selected 'working' correlation structures are: stationary MA(1) (moving average order 1), stationary EQC (equi-correlations), and independence (I). Note that the proposed dynamic model (4.4) produces time dependent covariates based non-stationary correlation structure (4.6), whereas the autocorrelations based existing SGEE approaches use stationary such as traditional AR(1), MA(1) and EQC structures. The repeated binary data were generated as in the last section. The simulated estimates using $K = 50$ and 100, for example are given in Tables 4.2 and 4.3, respectively. For longitudinal correlations, we choose its index as $\rho = 0.1, 0.5$. The estimates of the nonparametric function along with the true functions are displayed in Figure 4.1.

The SQL estimates of the function $\psi(\cdot)$ computed following (4.9) are displayed in Figure 4.1. This SQL approach appears to perform very well for the true quadratic

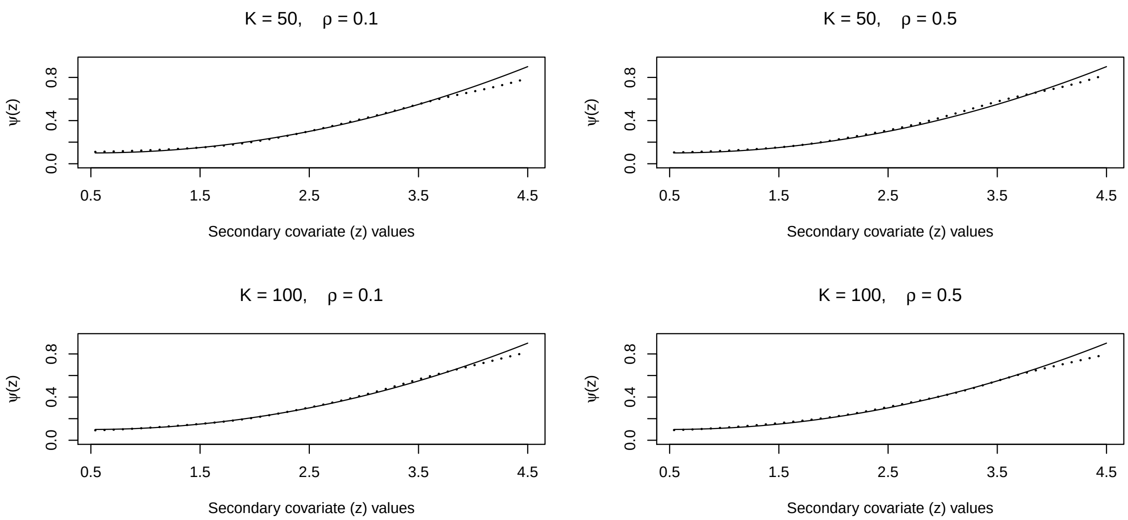

Figure 4.1: The plots for the true (thick curve) and estimated (dotted curve) nonparametric function $\psi(\cdot)$ for the SLDCP model based on $\boldsymbol{\beta}$ estimate produced by the SGQL approach. The bandwidth $b = K^{-1/5}$.

nonparametric function. Next, the simulation results in Table 4.2 and 4.3 show that the proposed SGQL approach appears to produce regression estimates with smaller MSE (mean squared error) than other SGEE approaches including the SGEE(I) approach, indicating its superiority. For example, for large $\rho = 0.5$, the results for $K = 50$ in Table 4.2 show that $\beta_1 = 0.5$ and $\beta_2 = 0.5$ are estimated by the SGQL approach with MSEs 0.2845 and 1.1146, while the SGEE(EQC) produces the estimates with larger MSEs 0.3154 and 1.2638, respectively; the independence approach SGEE(I) produces the estimates with MSEs 0.4287 and 1.8196, respectively, which are the worst performance. Furthermore, as expected, all approaches produce the regression estimates with similar MSEs when ρ is small, that is, $\rho = 0.1$. Next, when the results from Table 4.3 for $K = 100$ are compared with those in Table 4.2 for $K = 50$, the larger cluster number appears to produce estimates with smaller MSEs, as expected. The SMM approach explained in Section 4.1.2.3 also appears to produce estimates of ρ close to its true value. Thus the proposed SGQL approach performs well

ρ	Methods	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\rho}$
0.1	SGQL	SM	0.5531	0.5611		0.0925
		SSE	0.5170	1.1062		0.0780
		MSE	0.2698	1.2261		
	SGEE(MA(1))	SM	0.5529	0.5555	0.0975	
		SSE	0.5175	1.0933	0.0821	
		MSE	0.2703	1.1973		
	SGEE(EQC)	SM	0.5551	0.5610	0.0469	
		SSE	0.5203	1.0953	0.0657	
		MSE	0.2735	1.2022		
	SGEE(I)	SM	0.5504	0.5498		
		SSE	0.5218	1.1091		
		MSE	0.2746	1.2313		
0.5	SGQL	SM	0.5529	0.5115		0.4517
		SSE	0.5310	1.0562		0.0864
		MSE	0.2845	1.1146		
	SGEE(MA(1))	SM	0.5701	0.4680	0.5249	
		SSE	0.5955	1.2323	0.0736	
		MSE	0.3592	1.5181		
	SGEE(EQC)	SM	0.5645	0.5575	0.3747	
		SSE	0.5582	1.1233	0.0855	
		MSE	0.3154	1.2638		
	SGEE(I)	SM	0.5409	0.5234		
		SSE	0.6538	1.3494		
		MSE	0.4287	1.8196		

Table 4.2: Simulated means (SMs), simulated standard errors (SSEs) and mean square errors (MSEs) of the SGQL and SGEE estimates of the regression parameter $\beta_1 = 0.5$ and $\beta_2 = 0.5$, under LDCP AR(1) correlation model for selected values of correlation index parameter ρ with $K = 50$, $n = 4$, and 1000 simulations. The covariates $\boldsymbol{x}_{ij}$'s are given by (4.31). The bandwidth $b = K^{-1/5}$.

in estimating the function $\psi(\cdot)$, main regression parameters and the correlation index parameter involved in the SLDCP model (4.4). This SGQL approach also performs better than any selected SGEE approaches.

ρ	Methods	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\alpha}$	$\hat{\rho}$
0.1	SGQL	SM	0.5235	0.5237		0.0956
		SSE	0.3587	0.7783		0.0557
		MSE	0.1291	0.6058		
	SGEE(MA(1))	SM	0.5232	0.5239	0.1009	
		SSE	0.3588	0.7788	0.0587	
		MSE	0.1292	0.6065		
	SGEE(EQC)	SM	0.5246	0.5275	0.0509	
		SSE	0.3602	0.7790	0.0455	
		MSE	0.1302	0.6070		
	SGEE(I)	SM	0.5237	0.5335		
		SSE	0.3622	0.7828		
		MSE	0.1316	0.6132		
0.5	SGQL	SM	0.5344	0.5355		0.4794
		SSE	0.3652	0.7521		0.0564
		MSE	0.1344	0.5663		
	SGEE(MA(1))	SM	0.5361	0.5196	0.5290	
		SSE	0.4019	0.8712	0.0515	
		MSE	0.1627	0.7586		
	SGEE(EQC)	SM	0.5401	0.5535	0.3761	
		SSE	0.3848	0.8077	0.0570	
		MSE	0.1495	0.6546		
	SGEE(I)	SM	0.5338	0.5593		
		SSE	0.4491	0.9704		
		MSE	0.2026	0.9442		

Table 4.3: Simulated means (SMs), simulated standard errors (SSEs) and mean square errors (MSEs) of the SGQL and SGEE estimates of the regression parameter $\beta_1 = 0.5$ and $\beta_2 = 0.5$, under LDCP AR(1) correlation model for selected values of correlation index parameter ρ with $K = 100$, $n = 4$, and 1000 simulations. The covariates $\boldsymbol{x}_{ij}$'s are given by (3.105). The bandwidth $b = K^{-1/5}$.

4.1.5 An illustration: Fitting the SLDCP model to the longitudinal infectious disease data

To illustrate the proposed semi-parametric LDCP (4.4) with a real life data, in this section we reanalyze the respiratory infection (0 =no, 1 =yes) data earlier studied by some authors (Zeger and Karim, 1991, Diggle et al., 1994, Lin and Carroll, 2001).

These binary data for the presence of respiratory infection were collected from 275 preschool-age children examined every quarter for up to six consecutive quarters. In our notation, y_{ij} indicates the infection status of the ith ($i = 1, \ldots, 275$) child collected on jth ($j = 1, \ldots, n_i$) quarter with max $n_i = 6$. A variety of primary covariates, namely, vitamin A deficiency, sex, height, and stunting status; and a secondary covaraite, namely the age of the child in unit of month, were recorded. In our notation, these primary and secondary covariates are denoted by $\boldsymbol{x}_{ij}(j)$ and z_{ij} respectively. Similar to the earlier studies, it is of main interest to find the effects ($\boldsymbol{\beta}$) of the primary covariates while fitting the secondary covariates through a nonparametric function $\psi(z_{ij})$. For the purpose, Lin and Carroll (2001, Section 8) for example, fitted a semi-parametric marginal model with binary means

$$\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) = \frac{\exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij}))}{1 + \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij}))},$$

as in (4.3). As far as the correlation model is concerned, they did not model the correlations of the repeated binary responses. Instead they used the so-called 'working' UNS correlation structure (4.33) based SGEE approach for efficient estimation of the nonparametric function and other parameters as well. However, as it was demonstrated in Section 4.1.4.1, this UNS based SGEE approach (SGEE(UNS)) turned to be undesirable as it produced less efficient regression estimates under the SLDCP model (4.4) (see also Sutradhar and Das, 1999). Furthermore, because the SGQL approach discussed in Section 4.1.4.2 performs very well in estimating the parameters, we fitted the SLDCP model using this SGQL estimation approach. The nonparametric function estimates $\hat{\psi}(\text{age})$ using independence assumption based SQL approach are displayed in Figure 4.2. These estimates, unlike in Lin and Carroll (2001), in general show a linear negative effect of age rather than any quadratic effect. The estimates of the regression effects of the primary covariates involved in the parametric regression function obtained by using the SGQL approach are shown in Table 4.4. To construct a confidence interval for the estimated age effect, one may use sandwich method to estimate its pointwise standard errors (Lin and Carroll, 2001). For estimating the standard error of the estimator of the correlation index parameter ρ, we recommend to generate a large size of data according to the model (4.3) - (4.4) with the estimated parameters and nonparametric function, then calculate the sample standard error of the estimator (4.18). However, the secondary covariate and the correlation index parameter are not our primary interest, so we do not include these estimates here.

In a longitudinal study, the mean and variance of the data usually change with regard to time, mainly due to the influence of time dependent primary and secondary covariates; because of this, it may not be enough to examine only the effects of the primary covariates in such a study. For this reason, we computed the averages of the

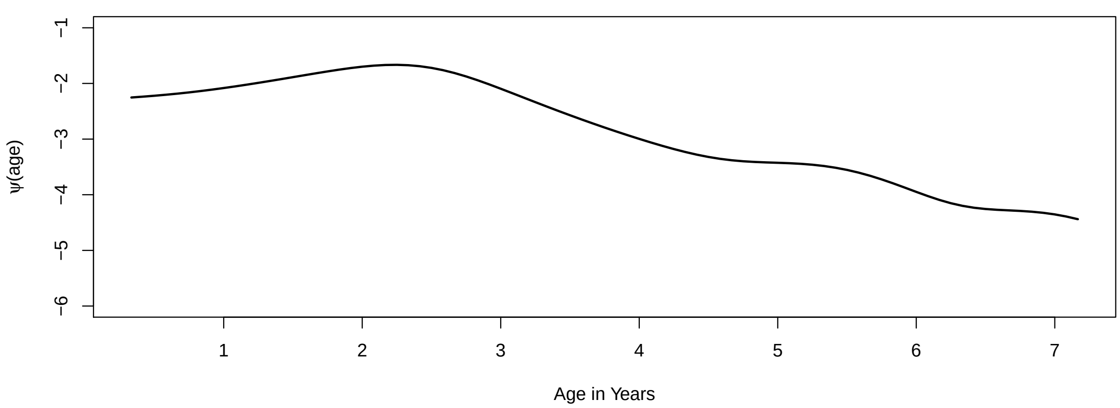

Figure 4.2: Estimated $\psi(\cdot)$ function for the unbalanced infectious disease data using the semi-parametric LDCP (SLDCP) model. The bandwidth $b = \left(\frac{\text{age range}}{4}\right) K^{-1/5}$, where K is the number of individuals in the relevant data.

Primary covariates	SLDCP Estimate	SLDCP SE	PLDCP Estimate	PLDCP SE
Vitamin A deficiency	0.576	0.448	0.701	0.443
Seasonal Cosine	-0.579	0.170	-0.569	0.167
Seasonal Sine	-0.156	0.168	-0.165	0.168
Sex	-0.515	0.227	-0.399	0.224
Height	-0.027	0.025	-0.044	0.025
Stunting	0.464	0.407	0.168	0.398
Age as a primary covariate	–	–	-0.388	0.078
Intercept	–	–	-1.277	0.259
correlation index parameter ρ	0.020	–	0.028	–

Table 4.4: Primary regression effect estimates along with their standard errors for the respiratory infectious data under the semi-parametric LDCP (SLDCP) (4.4) and fully parametric LDCP (PLDCP) models. The bandwidth $b = \left(\frac{\text{age range}}{4}\right) K^{-1/5}$, where K is the number of individuals.

binary data along with their estimated means over the time range under the SLDCP model. For a given time j, these averages are

$$\begin{aligned} \bar{y}_j &= \frac{\sum_{i=1}^{K} y_{ij}}{K} \\ \hat{\mu}_j(\text{for SLDCP model}) &= \frac{\sum_{i=1}^{K} \hat{\mu}_{ij}(\hat{\boldsymbol{\beta}}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}, z_{ij}))}{K}, \end{aligned} \tag{4.35}$$

respectively. These averages in Figure 4.3 show that the fitted means under the model are somehow close to the mean functions of the binary observations (in solid green). In particular, the observed and fitted means seem to show the same pattern. Furthermore, because the estimated nonparametric functions in Figure 4.2 show a negative linear effect of age on the responses, we have also fitted a parametric LDCP (PLDCP) model by treating age as an additional primary covariate. We remark that in other problems in practice, one may obtain a complicated pattern for the function $\psi(\cdot)$. Moreover, as the Table 4.4 shows, the estimated effect values of some of the

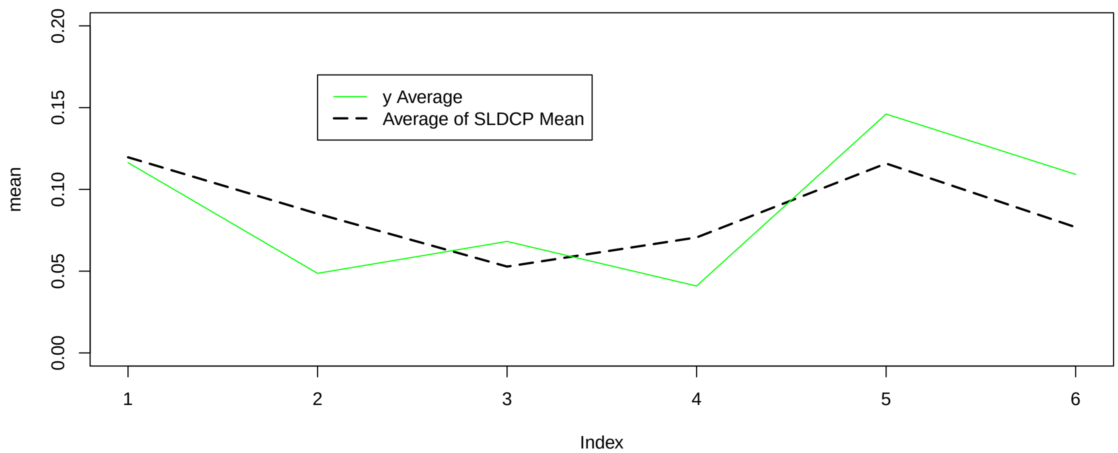

Figure 4.3: The average of the estimated means under the SLDCP model and the average of y values at each longitudinal index (time) point for the unbalanced infectious disease data. The bandwidth $b = \left(\frac{\text{age range}}{4}\right) K^{-1/5}$, where K is the number of individuals in the relevant data.

covariates such as sex and stunting are quite different under this PLDCP model as compared to the semiparametric LDCP (SLDCP) model. This difference may stem from the inclusion of age effect in the parametric function, since Fig. 4.2 shows that the detailed effect of age is not totally linear.

In summary, we now choose to interpret the effects of the primary covariates under the SLDCP model as opposed to the parametric LDCP model. To be specific, the Vitamin A deficiency (yes/no) has a large positive effect 0.57 on the probability of having respiratory infection in a child. The negative value -0.52 for the sex effect shows that female child (coded as 1) has smaller probability of having respiratory infection. As far as the nonparametric function effect is concerned, the estimated function under the SLDCP model shows that as age increases the infection probability decreases as one may expect.

4.2 SBDL (semi-parametric binary dynamic logit) model for longitudinal binary data

Sutradhar (2011) considered a BDL model for longitudinal binary data given by

$$\begin{aligned} Pr[Y_{i1} = 1|\boldsymbol{x}_{i1}] &= \pi_{i1}(\boldsymbol{\beta}|\boldsymbol{x}_{i1}) = \frac{\exp(\boldsymbol{x}_{i1}^\top \boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_{i1}^\top \boldsymbol{\beta})}, \quad \text{and} \qquad (4.36) \\ Pr[Y_{ij} = 1|y_{i,j-1}, \boldsymbol{x}_{ij}] &= \frac{\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \theta y_{i,j-1})}{1+\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \theta y_{i,j-1})} \text{ for } j = 2, \ldots, n_i, \\ &= p_{i,j|j-1}(\boldsymbol{\beta}, \theta|\boldsymbol{x}_{ij}, y_{i,j-1}), \qquad (4.37) \end{aligned}$$

where θ is a dynamic dependence parameter which is quite different than ρ in the LDCP model (4.2). More specifically, θ parameter in (4.37) can range from $-\infty$ to $+\infty$, whereas the ρ parameter in (4.2) must satisfy a range restriction so that the conditional probability $\lambda_{i,j|j-1}(\cdot)$ may range from 0 to 1. Furthermore, the marginal mean (and hence variance) at a given time point under the LDCP model (4.1)$-$(4.2) depends on the covariates at that time point only, whereas the marginal mean (and hence variance) at a given time point under the BDL model (4.36)$-$(4.37) is a function of the covariate history up to the present time point, thus, generating a recursive relationship among the means. To be specific, at time j, the LDCP model (4.1)$-$(4.2) has the marginal means stated by (4.1), whereas the BDL model (4.36)$-$(4.37) produce the corresponding marginal means as

$$\pi_{ij}(\boldsymbol{\beta}, \theta|\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{ij}) = \pi_{ij}^* + \pi_{i,j-1}(\boldsymbol{\beta}, \theta|\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{i,j-1}) \left[\tilde{\pi}_{ij} - \pi_{ij}^*\right], \qquad (4.38)$$

for $j = 2, \ldots, n_i$, where

$$\tilde{\pi}_{ij} = \frac{\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \theta)}{1+\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \theta)}, \text{ and } \pi_{ij}^* = \frac{\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})}{1+\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\beta})},$$

with $\pi_{i1}(\cdot) = \mu_{i1}(\boldsymbol{\beta}|\boldsymbol{x}_{i1}) = \pi^*_{i1}$. These marginal means in (4.38) show a recursive relationship. For further details on the basic properties including the correlations among repeated responses under these two models, see Sutradhar (2011, Sections 7.4, 7.7.2) and Sutradhar and Farrell (2007), for example.

Note that many longitudinal binary data in socio-economic and bio-medical fields appear to follow the marginal means pattern (4.38) as compared to (4.1). For example, in a socio-economic problem, the unemployment/employment status of an individual at a given year is likely to be a function of all mean employment status from all past years. Similarly, in a bio-medical field, the asthma status of an individual at a given month or year would likely be the function of all past asthma status of the individual. For this reason, in this section we concentrate our attention to the longitudinal binary data satisfying the BDL ((4.36)−(4.37)) type model that produces the recursive marginal means given by (4.38). Further note that the BDL model (4.36)−(4.37) is written in terms of the primary covariates $\{\boldsymbol{x}_{ij}, j = 1, \ldots, n_i\}$ only. As a main purpose of this thesis, we now generalize the BDL model to the semi-parametric setup by considering secondary covariates denoted earlier by $\{z_{ij}(t_{ij}), j = 1, \ldots, n_i\}$, and their effects on the binary responses accommodated nonparametrically by a smooth function $\psi(z_{ij})$. One may then extend the BDL model (4.36)−(4.37) to the longitudinal semi-parametric setup and write a semi-parametric BDL (SBDL) model as

$$
\begin{aligned}
Pr[Y_{i1} = 1|\boldsymbol{x}_{i1}, z_{i1}] &= \frac{\exp(\boldsymbol{x}_{i1}^\top\boldsymbol{\beta} + \psi(z_{i1}))}{1 + \exp(\boldsymbol{x}_{i1}^\top\boldsymbol{\beta} + \psi(z_{i1}))} \\
&= \pi_{i1}(\boldsymbol{\beta}, \psi(z_{i1})|\boldsymbol{x}_{i1}, z_{i1}), \quad \text{and} \qquad (4.39) \\
Pr[Y_{ij} = 1|y_{i,j-1}, \boldsymbol{x}_{ij}, z_{ij}] &= \frac{\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \theta y_{i,j-1} + \psi(z_{ij}))}{1 + \exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \theta y_{i,j-1} + \psi(z_{ij}))} \text{ for } j = 2, \ldots, n_i, \\
&= p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij})|\boldsymbol{x}_{ij}, z_{ij}, y_{i,j-1}). \qquad (4.40)
\end{aligned}
$$

Notice that unlike the BDL model (4.36)−(4.37), the main regression parameter $\boldsymbol{\beta}$

and the dynamic dependence (or correlation index) parameter θ in the SBDL model (4.39)$-$(4.40) can not be consistently estimated unless the function $\psi(\cdot)$ is estimated. More specifically, if the presence of $\psi(\cdot)$ is ignored and $\boldsymbol{\beta}$ and θ are jointly estimated using the data that follow the model (4.39)$-$(4.40), one would then obtain biased and hence mean squared error inconsistent estimates for these parameters. The main objective of this section is to obtain a consistent estimator $\hat{\psi}(z_{ij}|\boldsymbol{\beta}, \theta)$ for $\psi(z_{ij})$ assuming that $\boldsymbol{\beta}$ and θ are known and then estimate $\boldsymbol{\beta}$ and θ jointly by exploiting the modified SBDL model given by

$$\bar{\pi}_{i1}(\boldsymbol{\beta}, \hat{\psi}(\boldsymbol{\beta}, z_{i1})|\boldsymbol{x}_{i1}, z_{i1}) = \frac{\exp(\boldsymbol{x}_{i1}^{\top}\boldsymbol{\beta} + \hat{\psi}(\boldsymbol{\beta}, z_{i1}))}{1 + \exp(\boldsymbol{x}_{i1}^{\top}\boldsymbol{\beta} + \hat{\psi}(\boldsymbol{\beta}, z_{i1}))}, \tag{4.41}$$

$$\begin{aligned} & \bar{p}_{i,j|j-1}(\boldsymbol{\beta}, \theta, \hat{\psi}(\boldsymbol{\beta}, \theta, z_{ij})|\boldsymbol{x}_{ij}, z_{ij}, y_{i,j-1}) \\ = \; & \frac{\exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \theta y_{i,j-1} + \hat{\psi}(\boldsymbol{\beta}, \theta, z_{ij}))}{1 + \exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\beta} + \theta y_{i,j-1} + \hat{\psi}(\boldsymbol{\beta}, \theta, z_{ij}))}, \end{aligned} \tag{4.42}$$

for $j = 2, \ldots, n_i$. The estimation via a semi-parametric conditional quasi-likelihood (SCQL) approach for the estimation of the function $\psi(\cdot)$, and a joint MLE (maximum likelihood) approach for the estimation of $\boldsymbol{\beta}$ and θ is presented in Section 4.2.2. Because one of the purpose of the estimation of the model is to understand the data through estimation of the basic properties such as mean, variance and correlations of the repeated binary responses, we first provide these basic properties of the SBDL model (4.39)$-$(4.40) in Section 4.2.1. The consistency of all estimators is shown in Section 4.2.3. We also discuss the finite sample properties of the estimators through a simulation study in Section 4.2.4. The proposed longitudinal semi-parametric model and the estimation methodology are then illustrated by reanalyzing the well known respiratory infection status data earlier analyzed by some authors such as Zeger and Karim (1991), Diggle et al. (1994), Lin and Carroll (2001) (see also Sutradhar et al., 2016). This is done in Section 4.2.5.

4.2.1 Basic properties of the SBDL model

The proposed semi-parametric BDL (SBDL) model is stated in (4.39)−(4.40), which similarly to (4.36)−(4.38) produces the recursive means as

$$\pi_{ij}(\boldsymbol{\beta},\theta,\psi(\cdot)) = E[Y_{ij}] = Pr[Y_{ij}=1]$$
$$= \begin{cases} \pi_{i1}(\boldsymbol{\beta},\psi(z_{i1})) & j=1 \\ \pi^*_{ij}(\boldsymbol{\beta},\psi(z_{ij})) + \pi_{i,j-1}(\boldsymbol{\beta},\theta,\psi(\cdot))[\tilde{\pi}_{ij}(\boldsymbol{\beta},\theta,\psi(z_{ij})) - \pi^*_{ij}(\boldsymbol{\beta},\psi(z_{ij}))] & j=2,\ldots,n_i, \end{cases} \tag{4.43}$$

where

$$\begin{aligned} \tilde{\pi}_{ij}(\boldsymbol{\beta},\theta,\psi(z_{ij})) &= \frac{\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\theta+\psi(z_{ij}))}{1+\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\theta+\psi(z_{ij}))}, \text{ and} \\ \pi^*_{ij}(\boldsymbol{\beta},\psi(z_{ij})) &= \frac{\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\psi(z_{ij}))}{1+\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\beta}+\psi(z_{ij}))}, \text{ satisfying } \pi^*_{i1}(\boldsymbol{\beta},\psi(z_{i1})) = \pi_{i1}(\boldsymbol{\beta},\psi(z_{i1})). \end{aligned}$$

It is obvious that the variances are given by

$$\sigma_{ijj}(\boldsymbol{\beta},\theta,\psi(\cdot)) = \text{var}[Y_{ij}] = \pi_{ij}(\boldsymbol{\beta},\theta,\psi(\cdot))[1-\pi_{ij}(\boldsymbol{\beta},\theta,\psi(\cdot))], \text{ for } j=1,\ldots,n_i. \tag{4.44}$$

As far as the correlation properties of the SBDL model ((4.39)−(4.40)) is concerned, for $j < k$, following Sutradhar and Farrell (2007), for example, one may compute the pair-wise covariances as

$$\begin{aligned} \sigma_{ijk}(\boldsymbol{\beta},\theta,\psi(\cdot)) &= \text{Cov}(Y_{ij},Y_{ik}) \\ &= \pi_{ij}(\boldsymbol{\beta},\theta,\psi(\cdot))[1-\pi_{ij}(\boldsymbol{\beta},\theta,\psi(\cdot))] \\ &\times \Pi_{u=j+1}^k[\tilde{\pi}_{iu}(\boldsymbol{\beta},\theta,\psi(z_{iu})) - \pi^*_{iu}(\boldsymbol{\beta},\psi(z_{iu}))], \end{aligned} \tag{4.45}$$

yielding the pair-wise lag $(k-j)$ correlations as

$$\begin{aligned}\text{Corr}(Y_{ij}, Y_{ik}) &= \sqrt{\frac{\sigma_{ijj}(\boldsymbol{\beta}, \theta, \psi(\cdot))}{\sigma_{ikk}(\boldsymbol{\beta}, \theta, \psi(\cdot))}} \\ &\times \ \Pi_{u=j+1}^{k}[\tilde{\pi}_{iu}(\boldsymbol{\beta}, \theta, \psi(z_{iu})) - \pi^*_{iu}(\boldsymbol{\beta}, \psi(z_{iu}))] \end{aligned} \tag{4.46}$$

which satisfies the full range from -1 to 1, as

$$0 < \tilde{\pi}_{iu}(\boldsymbol{\beta}, \theta, \psi(z_{iu})), \pi^*_{iu}(\boldsymbol{\beta}, \psi(z_{iu})) < 1.$$

We remark that understanding the basic properties of the data requires the estimation of the nonparametric functions as well as the parameters $\boldsymbol{\beta}$ and θ involved in the formulas (4.43), (4.44), and (4.46). We deal with this estimation issue in the next section.

4.2.2 SBDL model fitting

Fitting the SBDL model (4.39)−(4.40) to the repeated binary data requires the estimation of the nonparametric function $\psi(\cdot)$, and the model parameters $\boldsymbol{\beta}$ and θ, where $\boldsymbol{\beta}$ is the main regression effects and θ is the dynamic dependence or correlation index parameters. We provide their step by step consistent estimation as follows.

4.2.2.1 SCQL estimation of $\psi(\cdot)$ under the SBDL model

Note that it is of primary interest to estimate the regression effects $\boldsymbol{\beta}$ and the dynamic dependence parameter θ involved in the SBDL model (4.40) consistently and as efficiently as possible. However, one can not obtain the consistent estimators of $\boldsymbol{\beta}$ and θ in (4.40) without consistently estimating the nonparametric function $\psi(z_{ij})$. Thus, for known $\boldsymbol{\beta}$ and θ, we first develop a consistent estimator $\hat{\psi}(\boldsymbol{\beta}, z_{ij})$ for the function

$\psi(z_{ij})$.

It follows from the SBDL model (4.40) that a consistent estimator of $\psi(\cdot)$ can be obtained by exploiting only the conditional means and variances of the repeated responses $\{y_{ij}, j = 1, \ldots, n_i\}$. Observe from the model that $p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij})|\boldsymbol{x}_{ij}, z_{ij}, y_{i,j-1})$ is the conditional mean (probabilities) of y_{ij} for $j = 2, \ldots, n_i$. For technical convenience, for $j = 1$, define $y_{i0} = 0$. It then follows by (4.39) and (4.40) that

$$\pi_{i1}(\boldsymbol{\beta}, \psi(z_{i1})) = p_{i,1|0}(\boldsymbol{\beta}, \theta, \psi(z_{i1})|\boldsymbol{x}_{i1}, z_{i1}, y_{i0} = 0).$$

Thus, in general, conditioning on the past response, we may now write the conditional means and variances of y_{ij} as

$$\begin{aligned}
E[Y_{ij}|y_{i,j-1}, \boldsymbol{x}_{ij}, z_{ij}] &= p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij})|\boldsymbol{x}_{ij}, z_{ij}, y_{i,j-1}) \\
\mathrm{var}[Y_{ij}|y_{i,j-1}, \boldsymbol{x}_{ij}, z_{ij}] &= p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij})|\boldsymbol{x}_{ij}, z_{ij}, y_{i,j-1}) \\
&\times [1 - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij})|\boldsymbol{x}_{ij}, z_{ij}, y_{i,j-1})], \qquad (4.47)
\end{aligned}$$

for all $j = 1, \ldots, n_i$. These conditional means and variances in (4.47) will be exploited to write a QL (quasi-likelihood) (Wedderburn, 1974) estimating equation for the estimation of $\psi(z_{ij})$. We refer this approach as the semi-parametric conditional QL (SCQL) estimation approach.

For this estimation purpose, without loss of generality, we use z_0 for z_{ij} for given i and j, and hence estimate $\psi(z_0)$ at all possible values of z_0 corresponding to all i and j. We remark that it is impossible to estimate $\boldsymbol{\beta}$ and θ consistently without estimating $\psi(z_0)$ consistently. Thus, for known $\boldsymbol{\beta}$ and θ, following the QL approach of Wedderburn (1974), one may now solve the SCQL estimating equation

$$\begin{aligned} & \sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\ \frac{\partial p_{i,j|j-1}(\boldsymbol{\beta},\theta,\psi(z_0))}{\partial \psi(z_0)} \left(\frac{y_{ij} - p_{i,j|j-1}(\boldsymbol{\beta},\theta,\psi(z_0))}{p_{i,j|j-1}(\boldsymbol{\beta},\theta,\psi(z_0))\{1 - p_{i,j|j-1}(\boldsymbol{\beta},\theta,\psi(z_0))\}} \right) \\ = \ & \sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{ y_{ij} - p_{i,j|j-1}(\boldsymbol{\beta},\theta,\psi(z_0))\right\} = g(\psi(z_0),\boldsymbol{\beta},\theta) = 0, \end{aligned} \tag{4.48}$$

to obtain a consistent estimate of $\psi(z_0)$ (see also Severini and Staniswalis, 1994). In (4.48), $w_{ij}(z_0)$ is a kernel weight defined as

$$w_{ij}(z_0) = p_{ij}(\frac{z_0 - z_{ij}}{b})/\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(\frac{z_0 - z_{ij}}{b}) \tag{4.49}$$

with p_{ij} being the kernel density, as discussed in (3.12)$-$(3.14). Note that for $w_{ij}(z_0) = 1$, the SCQL estimating equation (4.48) reduces to the well known QL (quasi-likelihood) equation. Here b is a suitable bandwidth parameter. We assume that this parameter is chosen such that the bias and variance of the estimator of the function $\psi(z_{ij})$ will be minimum. By this token, b may be optimally chosen as $b \propto K^{-1/5}$ (Pagan and Ullah, 1999, Altman, 1990). More specifically, as we explained in Section 4.1.4, one may use $b = c_0 K^{-1/5}$ where the constant c_0 can be estimated, for example, following Horowitz (2009, Section 2.7) (see also Powell and Stoker, 1996).

Now for known $\boldsymbol{\beta}$ and θ, we may solve the SCQL estimating equation (4.48) by using the iterative equation given by

$$\begin{aligned} \hat{\psi}(z_0,\boldsymbol{\beta},\theta)_{(r+1)} \ &= \ \hat{\psi}(z_0,\boldsymbol{\beta},\theta)_{(r)} \\ &- \ \left[\{g'_{\psi(z_0)}(\psi(z_0),\boldsymbol{\beta},\theta)\}^{-1} g(\psi(z_0),\boldsymbol{\beta},\theta)\right]_{|\psi(z_0)=\hat{\psi}(z_0,\boldsymbol{\beta},\theta)_{(r)}}, \end{aligned} \tag{4.50}$$

where (r) indicates the rth iteration, and $g'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta}, \theta)$ has the formula

$$\begin{aligned} & g'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta}, \theta) \\ = \quad & -\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\{1 - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\}]. \end{aligned} \tag{4.51}$$

4.2.2.2 Joint estimation: Semi-parametric maximum likelihood (SML) estimation of $\boldsymbol{\beta}$ and θ

For the SBDL model (4.39)−(4.40), $\psi(z_{ij})$ is estimated by solving the semi-parametric conditional QL (SCQL) estimating equation given by (4.48). For given $\boldsymbol{\beta}$ and θ, this estimator is denoted by $\hat{\psi}(z_{ij}, \boldsymbol{\beta}, \theta)$ as in (4.50). Because the regression and the dynamic dependence parameters in the SBDL model appear in the conditional mean functions in a similar way, it is convenient to estimate them jointly. Let $\boldsymbol{\phi} = (\boldsymbol{\beta}^\top, \theta)^\top$.

Now by using $\hat{\psi}(z_{ij}, \boldsymbol{\beta}, \theta)$ from (4.50) for the true $\psi(\cdot)$, one may re-express the marginal and conditional probabilities from (4.39)−(4.40) as

$$\begin{aligned} \bar{\pi}_{i1}(\boldsymbol{\beta}, \hat{\psi}(z_{i1}, \boldsymbol{\beta})) &= \frac{\exp(\boldsymbol{x}_{i1}^\top \boldsymbol{\beta} + \hat{\psi}(z_{i1}, \boldsymbol{\beta}))}{1 + \exp(\boldsymbol{x}_{i1}^\top \boldsymbol{\beta} + \hat{\psi}(z_{i1}, \boldsymbol{\beta}))}, \text{ and} \\ \bar{p}_{ij|j-1}(\boldsymbol{\beta}, \theta, \hat{\psi}(z_{ij}, \boldsymbol{\phi})) &= \frac{\exp[\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + y_{i,j-1}\theta + \hat{\psi}(z_{ij}, \boldsymbol{\phi})]}{1 + \exp[\boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + y_{i,j-1}\theta + \hat{\psi}(z_{ij}, \boldsymbol{\phi})]}, \end{aligned}$$

respectively, and write the likelihood function for $\boldsymbol{\phi} = (\boldsymbol{\beta}^\top, \theta)^\top$ as

$$\begin{aligned} L(\boldsymbol{\beta}, \theta, \hat{\psi}(\cdot, \boldsymbol{\beta}, \theta)) = \Pi_{i=1}^{K} \Big[& \{\bar{\pi}_{i1}(\boldsymbol{\beta}, \hat{\psi}(z_{i1}, \boldsymbol{\beta}))\}^{y_{i1}} \{1 - \bar{\pi}_{i1}(\boldsymbol{\beta}, \hat{\psi}(z_{i1}, \boldsymbol{\beta}))\}^{1-y_{i1}} \\ \times \Pi_{j=2}^{n_i} & \{\bar{p}_{i,j|j-1}(\boldsymbol{\beta}, \theta, \hat{\psi}(z_{ij}, \boldsymbol{\beta}, \theta)|y_{i,j-1})\}^{y_{ij}} \{1 - \bar{p}_{i,j|j-1}(\boldsymbol{\beta}, \theta, \hat{\psi}(z_{ij}, \boldsymbol{\beta}, \theta)|y_{i,j-1})\}^{1-y_{ij}} \Big], \end{aligned} \tag{4.52}$$

leading to the log likelihood estimating equation for $\boldsymbol{\phi}$ given by

$$\begin{aligned}
\boldsymbol{H}_K &= \frac{\partial \log L}{\partial \boldsymbol{\phi}} \\
&= \sum_{i=1}^{K}\sum_{j=1}^{n_i} y_{ij}\left[\begin{pmatrix} \boldsymbol{x}_{ij} \\ y_{i,j-1} \end{pmatrix} + \frac{\partial \hat{\psi}(z_{ij},\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right] - \sum_{i=1}^{K}\sum_{j=1}^{n_i} \bar{p}_{ij|j-1}\left[\begin{pmatrix} \boldsymbol{x}_{ij} \\ y_{i,j-1} \end{pmatrix} + \frac{\partial \hat{\psi}(z_{ij},\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right] \\
&= \sum_{i=1}^{K}\sum_{j=1}^{n_i} \left(y_{ij} - \bar{p}_{ij|j-1}\right)\left[\begin{pmatrix} \boldsymbol{x}_{ij} \\ y_{i,j-1} \end{pmatrix} + \frac{\partial \hat{\psi}(z_{ij},\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right] \; = \; 0, \qquad (4.53)
\end{aligned}$$

where we have used $y_{i0} = 0$ as a conventional notation. Notice that the likelihood equation (4.53) contains the derivative function $\frac{\partial \hat{\psi}(z_{ij},\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}$ which must be computed from the SCQL estimating equation (4.48) for $\psi(z_{ij})$ satisfying

$$\sum_{\ell=1}^{K}\sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij})\{y_{\ell u} - \bar{p}_{\ell,u|u-1}(\boldsymbol{\phi}, \hat{\psi}(z_{ij},\boldsymbol{\phi}))\} = 0.$$

This derivative function has the formula

$$\frac{\partial \hat{\psi}(z_{ij},\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = -\frac{\sum_{\ell=1}^{K}\sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij})\bar{p}_{\ell u|u-1}(z_{ij})\left[1 - \bar{p}_{\ell u|u-1}(z_{ij})\right]\begin{pmatrix} \boldsymbol{x}_{\ell u} \\ y_{\ell,u-1} \end{pmatrix}}{\sum_{\ell=1}^{K}\sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij})\bar{p}_{\ell u|u-1}(z_{ij})\left[1 - \bar{p}_{\ell u|u-1}(z_{ij})\right]}. \qquad (4.54)$$

Notice that this non-zero derivative (4.54) arises because of the use of the estimate of $\psi(\cdot)$ while estimating $\boldsymbol{\phi}$.

4.2.3 Asymptotic properties of the estimators of the SBDL model

4.2.3.1 Consistency of the nonparametric function estimator $\hat{\psi}(\cdot)$

Recall that under the SBDL model (4.40), $\hat{\psi}(z_0) \equiv \hat{\psi}(z_0, \boldsymbol{\beta}, \theta) \equiv \hat{\psi}(z_0, \boldsymbol{\phi})$ is the solution of the semi-parametric conditional quasi-likelihood (SCQL) estimating equation (4.48), that is, $g(\psi(z_0), \boldsymbol{\beta}, \theta) = 0$. By (4.48) and (4.51), a Taylor series expansion produces

$$
\begin{aligned}
& \hat{\psi}(z_0; \boldsymbol{\phi}) - \psi(z_0) \approx \frac{g(\psi(z_0), \boldsymbol{\beta}, \theta)}{g'_{\psi(z_0)}(\psi(z_0), \boldsymbol{\beta}, \theta)} \\
= & \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{y_{ij} - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\right\}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\{1 - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\}]} \\
= & \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{y_{ij} - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij}))\right\}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\{1 - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\}]} \\
& + \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\left\{p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij})) - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\right\}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\{1 - p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_0))\}]} \\
= & C_K + D_K\,, \qquad (4.55)
\end{aligned}
$$

where

$$
\begin{aligned}
C_K &= \frac{1}{G_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\left(y_{ij} - p_{i,j|j-1}\right) \text{ and} \\
D_K &= \frac{1}{G_K}\frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)\left(p_{i,j|j-1} - p_{i,j|j-1}(z_0)\right) \text{ with} \\
G_K &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(z_0)p_{i,j|j-1}(z_0)\left[1 - p_{i,j|j-1}(z_0)\right].
\end{aligned}
$$

Here $p_{ij}(z_0) \equiv p_{ij}(\frac{z_0 - z_{ij}}{b})$ is the kernel density defined in (3.13)−(3.14).

As in Lemma 4.2, it can be shown that

$$C_K = O_p(1/\sqrt{K}) \tag{4.56}$$

We now show that D_K approaches zero in order of $O(b^2)$.

Lemma 4.6. *The kernel density $p_{ij}(z_0)$ defined by (3.13)–(3.14) has the expectation given by*

$$\mathrm{E}\left[\sum_{j=1}^{n_i} p_{ij}(z_0)p_{ij|j-1}(z_0)\left[1 - p_{ij|j-1}(z_0)\right](z_{ij} - z_0)\middle|\boldsymbol{x}_i\right] \;=\; O(b^2), \tag{4.57}$$

where $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{in_i})^\top$.

Proof. Let $\boldsymbol{z}_i = (z_{i1}, \cdots, z_{in_i})^\top$ and $q_{ij} = \Pr(y_{ij} = 1|z_{i,j+1}, \boldsymbol{x}_i) = \Pr(y_{ij} = 1|\boldsymbol{x}_i)$ since the distribution of y_{ij} is independent of $z_{i,j+1}$ according to (4.39) and (4.40). Then $q_{ij} = \int \pi_{ij} f_i(\boldsymbol{z}_i|\boldsymbol{x}_i)d\boldsymbol{z}_i$, where π_{ij} is defined in (4.43), and $f_i(\boldsymbol{z}_i|\boldsymbol{x}_i)$ is the joint distribution of $\boldsymbol{z}_i$ conditional on $\boldsymbol{x}_i$. Also define

$$\begin{aligned}
g_j(z_{ij}; \boldsymbol{\beta}, \theta, z_0, \boldsymbol{x}_i) &= \mathrm{E}[p_{ij|j-1}(z_0)[1 - p_{ij|j-1}(z_0)]|z_{ij}, \mathbf{x}_i] \\
&= \sum_{y_{i,j-1}} p_{ij|j-1}(z_0)[1 - p_{ij|j-1}(z_0)]q_{i,j-1}^{y_{i,j-1}}\left(1 - q_{i,j-1}\right)^{1-y_{i,j-1}} \\
&= g_j(\boldsymbol{\beta}, \theta, z_0, \boldsymbol{x}_i)
\end{aligned}$$

because the conditional expectation is in fact independent of z_{ij}, and define $h_j(z_{ij}; \boldsymbol{x}_i)$ as the pdf of z_{ij} conditional on $\boldsymbol{x}_i$, then

$$\begin{aligned}
&\mathrm{E}\left[\sum_{j=1}^{n_i} p_{ij}(z_0)p_{ij|j-1}(z_0)\left[1 - p_{ij|j-1}(z_0)\right](z_{ij} - z_0)\middle|\boldsymbol{x}_i\right] \\
= &\sum_{j=1}^{n_i} \mathrm{E}_{z_{ij}}\left[p_{ij}(z_0)(z_{ij} - z_0)\mathrm{E}\{p_{ij|j-1}(z_0)[1 - p_{ij|j-1}(z_0)]|z_{ij}, \boldsymbol{x}_i\}\middle|\boldsymbol{x}_i\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n_i} \mathrm{E}_{z_{ij}} \left[p_{ij}(z_0)(z_{ij}-z_0) g_j(\boldsymbol{\beta}, \theta, z_0, \boldsymbol{x}_i) | \boldsymbol{x}_i \right] \\
&= \sum_{j=1}^{n_i} \int p_{ij}(z_0) g_j(\boldsymbol{\beta}, \theta, z_0, \boldsymbol{x}_i)(z_{ij}-z_0) h_j(z_{ij}; \boldsymbol{x}_i)\, dz_{ij}.
\end{aligned}
$$

Then as $h_j(z_{ij}; \boldsymbol{x}_i) = h_j(z_0; \boldsymbol{x}_i) + O(z_{ij} - z_0)$, it follows that

$$
\begin{aligned}
& \mathrm{E}\left[\sum_{j=1}^{n_i} p_{ij}(z_0) p_{ij|j-1}(z_0) \left[1 - p_{ij|j-1}(z_0)\right] (z_{ij}-z_0) \middle| \boldsymbol{x}_i \right] \\
&= \sum_{j=1}^{n_i} \int p_{ij}(z_0) \left[g_j(\boldsymbol{\beta}, \theta, z_0, \boldsymbol{x}_i) h_j(z_0; \boldsymbol{x}_i)(z_{ij}-z_0) + O\left((z_{ij}-z_0)^2\right) \right] dz_{ij} \\
&= \sum_{j=1}^{n_i} g_j(\boldsymbol{\beta}, \theta, z_0, \boldsymbol{x}_i) h_j(z_0; \boldsymbol{x}_i) \int p_{ij}(z_0)(z_{ij}-z_0)\, dz_{ij} + O(b^2) = O(b^2),
\end{aligned}
$$

because $p_{ij}(z_0)$ is symmetric about z_0 and $\int p_{ij}(z_0)\, O\left((z_{ij}-z_0)^2\right) dz_{ij}$ can be shown bounded in the order of $O(b^2)$. $\square$

Lemma 4.7. *The quantity D_K in (4.55) satisfies*

$$
D_K = O(b^2). \tag{4.58}
$$

Proof. By using

$$
p_{ij|j-1} - p_{ij|j-1}(z_0) = p_{ij|j-1}(z_0) \left[1 - p_{ij|j-1}(z_0)\right] \psi'(z_0)(z_{ij}-z_0) + O\left((z_{ij}-z_0)^2\right),
$$

we write

$$
\begin{aligned}
D_K &\approx \frac{\psi'(z_0)}{G_K} \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} p_{ij}(z_0) p_{ij|j-1}(z_0) \left[1 - p_{ij|j-1}(z_0)\right] (z_{ij}-z_0) \\
&= \frac{\psi'(z_0)}{G_K} \frac{1}{K} \sum_{i=1}^{K} \left\{ \sum_{j=1}^{n_i} p_{ij}(z_0) p_{ij|j-1}(z_0) \left[1 - p_{ij|j-1}(z_0)\right] (z_{ij}-z_0) \right.
\end{aligned}
$$

$$
- \mathrm{E}\left[\sum_{j=1}^{n_i} p_{ij}(z_0) p_{ij|j-1}(z_0)\left[1-p_{ij|j-1}(z_0)\right](z_{ij}-z_0)\middle|\boldsymbol{x}_i\right]\Bigg\}
$$

$$
+ \frac{\psi'(z_0)}{G_K}\frac{1}{K}\sum_{i=1}^{K}\mathrm{E}\left[\sum_{j=1}^{n_i} p_{ij}(z_0) p_{ij|j-1}(z_0)\left[1-p_{ij|j-1}(z_0)\right](z_{ij}-z_0)\middle|\boldsymbol{x}_i\right] = O(b^2),
$$

by law of large numbers (Breiman, 1992) and Lemma 4.6. For the first term, due to $p_{ij}(z_0)$, its variance is in the order of $O(b^2/K)$, so it is $O_p(b/\sqrt{K})$, which can be neglected. $\square$

By using (4.56) and Lemma 4.7 in (4.55), one obtains

$$
\hat{\psi}(z_0;\boldsymbol{\phi}) - \psi(z_0) = C_K + O(b^2) = O_p(1/\sqrt{K}) + O(b^2). \tag{4.59}
$$

It then follows that $\hat{\psi}(z_0;\boldsymbol{\beta},\theta)$ obtained from (4.48) is a $\sqrt{K}$-consistent estimator of $\psi(z_0)$ provided $Kb^4 \to 0$ for $K \to \infty$.

We remark that the consistency result in (4.59) was obtained by reducing the bias of the estimator $\hat{\psi}(\cdot)$. This consistency result holds when $Kb^4 \to 0$ as $K \to \infty$, implying that for a constant c^*, the bandwidth parameter b may be chosen as $b = c^* K^{-\alpha}$ with $\alpha > \frac{1}{4}$. A similar result with $\frac{1}{4} < \alpha \leq \frac{1}{3}$ is available in Lin and Carroll (2001), for example. It is however understandable that this choice for b value may not be optimal. This is because for any optimal selection one has to reduce both the bias and the variance of the estimator, whereas the convergence result in (4.59) was obtained by reducing the bias under the assumption that variance of the estimator would be finite based on the design covariates selection. A derivation for optimal choice for b value is beyond the scope of the present thesis. However, in the simulation study to be conducted in Section 4.2.4 and in further data analysis in Section 4.2.5, as indicated earlier, we consider an optimal value of $b = c_0 K^{-1/5}$ chosen under the independence setup (Pagan and Ullah, 1999; Horowitz, 2009, Section 2.7).

For c_0, we use $c_0 = \sigma_z$ (3.13), the standard deviation of the covariate z collected over the whole duration of the longitudinal study. Once again for the consistency of $\hat{\psi}(z_0; \boldsymbol{\phi})$ shown by (4.59) the mild condition $Kb^4 \to 0$ as $K \to \infty$ is sufficient.

4.2.3.2 Consistency of $\hat{\theta}$ and $\hat{\boldsymbol{\beta}}$

Let $\hat{\boldsymbol{\phi}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\theta})^\top$ denote the solution of the conditional maximum likelihood estimating equation (4.53) for $\boldsymbol{\phi}$. Then the asymptotic result of $\hat{\boldsymbol{\phi}}$ is given by the following lemma.

Lemma 4.8.

$$\begin{aligned}\sqrt{K}\left\{\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\} &= \boldsymbol{J}^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[\boldsymbol{W}_{1ij}(y_{i,j-1}) - \boldsymbol{W}_{2ij}\right](Y_{ij} - p_{ij|j-1}) \\ &+ O(\sqrt{Kb^4}) + o_p(1), \end{aligned} \tag{4.60}$$

where

$$\begin{aligned}\boldsymbol{W}_{1ij}(y_{i,j-1}) &= \begin{pmatrix} \boldsymbol{x}_{ij} \\ y_{i,j-1} \end{pmatrix} + \frac{\partial\hat{\psi}(z_{ij}, \boldsymbol{\phi})}{\partial\boldsymbol{\phi}}, \quad \textit{and} \\ \boldsymbol{W}_{2ij} &= \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}}\frac{1}{G_K(z_{i'j'})}\boldsymbol{W}_{1i'j'}p_{ij}(z_{i'j'}) \\ &\times \; p_{i'j'|j'-1}(\boldsymbol{\phi}, \psi(z_{i'j'}))\left[1 - p_{i'j'|j'-1}(\boldsymbol{\phi}, \psi(z_{i'j'}))\right],\end{aligned}$$

with $G_K(\cdot)$ defined as in (4.55).

Proof. A linear Taylor expansion of the left hand side of (4.53) about the true parameter value ϕ gives

$$\sqrt{K}\left\{\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right\} = \boldsymbol{J}_K^{-1}\left\{\sqrt{K}\,\boldsymbol{H}_K\right\} + o_p(1), \tag{4.61}$$

where

$$
\begin{aligned}
\boldsymbol{H}_K &= \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left(y_{ij}-\bar{p}_{ij|j-1}\right)\boldsymbol{W}_{1ij} \\
\boldsymbol{J}_K &= \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\bar{p}_{ij|j-1}\left(1-\bar{p}_{ij|j-1}\right)\left[\begin{pmatrix}\boldsymbol{x}_{ij}\\ y_{i,j-1}\end{pmatrix}+\frac{\partial\hat{\psi}(z_{ij};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right] \\
&\times \left[\begin{pmatrix}\boldsymbol{x}_{ij}\\ y_{i,j-1}\end{pmatrix}+\frac{\partial\hat{\psi}(z_{ij};\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right]^{\top}.
\end{aligned}
$$

Suppose that $\boldsymbol{J}=\lim_{K\to\infty}\boldsymbol{J}_K \ = \ \mathrm{E}_{\hat{\psi}(\cdot)}\boldsymbol{J}_K$. It then follows from (4.61) that

$$
\sqrt{K}\left\{\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}\right\}=\boldsymbol{J}^{-1}\left\{\sqrt{K}\,\boldsymbol{H}_K\right\}+o_p(1)\,. \tag{4.62}
$$

Now the lemma follows from (4.62) by writing

$$
\begin{aligned}
\boldsymbol{H}_K = \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}&\boldsymbol{W}_{1ij}\left[y_{ij}-p_{ij|j-1}(\boldsymbol{\phi},\psi(z_{ij}))\right] \\
-\ \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}&\boldsymbol{W}_{1ij}\left[\bar{p}_{ij|j-1}(\boldsymbol{\phi},\hat{\psi}(z_{ij};\boldsymbol{\phi}))-p_{ij|j-1}(\boldsymbol{\phi},\psi(z_{ij}))\right]
\end{aligned}
$$

and applying (4.59). □

Theorem 4.2. *The estimator $\hat{\boldsymbol{\phi}}$ (the solution of (4.53)) has the limiting (as $K\to\infty$) multivariate normal distribution given as*

$$
\sqrt{K}\left\{\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}-O(b^2)\right\} \ \ \to \ \ N(0,\boldsymbol{V}_{\boldsymbol{\phi}}), \tag{4.63}
$$

where

$$\boldsymbol{V}_{\boldsymbol{\phi}} = \boldsymbol{J}^{-1}\frac{1}{K}\left[\sum_{i=1}^{K}\mathrm{Var}\left\{\sum_{j=1}^{n_i}\left[\boldsymbol{W}_{1ij}(y_{i,j-1}) - \boldsymbol{W}_{2ij}\right](Y_{ij} - p_{ij|j-1})\right\}\right]\boldsymbol{J}^{-1}.$$

Proof. Because $\mathrm{E}[\boldsymbol{Y}_i - p_{ij|j-1}|y_{i,j-1}] = 0$, the theorem follows from Lemma 4.8 under some conditions (see Eqn. (3.76)), by applying Lindeberg-Feller central limit theorem (Bishop, Fienberg and Holland, 2007, Theorem 3.3.6) for independent random variables with non-identical distributions. □

Thus it follows from Lemma 4.8 and Theorem 4.2 that $\hat{\boldsymbol{\phi}}$ is $\sqrt{K}$ consistent estimator for $\boldsymbol{\phi}$, and has an asymptotic multivariate normal distribution provided $Kb^4 \to 0$ for $K \to \infty$.

4.2.4 A simulation study

The main objective of this simulation study is to examine the finite sample performance of the proposed SML (semi-parametric maximum likelihood) estimator of the regression parameter $\boldsymbol{\beta}$ and dynamic dependence parameter θ obtained by solving the SML estimating equation (4.53). Notice that the SML estimates for $\boldsymbol{\beta}$ and θ were obtained by using the SCQL (semi-parametric conditional quasi-likelihood) estimate for the function $\psi(\cdot)$. Thus the simulation study will also show the performance of this SCQL estimate for the nonparametric function.

We choose $K = 100$ individuals as a small sample size and $K = 300$ as a moderately large sample size. Next suppose that the ith individual provides $n_i = 4$ repeated binary responses for all $i = 1, \ldots, K$. The primary covariates are selected as:

For $K = 100,\ 300:$

$$x_{ij1}(j) = \begin{cases} \frac{1}{2} & \text{for } i = 1, \ldots, K/4 \text{ and } j = 1, 2 \\ 1 & \text{for } i = 1, \ldots, K/4 \text{ and } j = 3, 4 \\ \frac{-1}{2} & \text{for } i = K/4 + 1, \ldots, 3K/4 \text{ and } j = 1 \\ 0 & \text{for } i = K/4 + 1, \ldots, 3K/4 \text{ and } j = 2, 3 \\ \frac{1}{2} & \text{for } i = K/4 + 1, \ldots, 3K/4 \text{ and } j = 4 \\ \frac{j}{2n_i} & \text{for } i = 3K/4 + 1, \ldots, K \text{ and } j = 1, 2, 3, 4 \end{cases}$$
$$x_{ij2}(j) = \begin{cases} \frac{j-2.5}{2n_i} & \text{for } i = 1, \ldots, K/2 \text{ and } j = 1, 2, 3, 4 \\ 0 & \text{for } i = K/2 + 1, \ldots, K \text{ and } j = 1, 2 \\ \frac{1}{2} & \text{for } i = K/2 + 1, \ldots, K \text{ and } j = 3, 4. \end{cases} \tag{4.64}$$

Note that for $K = 100$, this design is the same as in (3.105).

As far as the secondary covariates (z_{ij}) and nonparametric functions ($\psi(z_{ij})$) are concerned, we choose them as in (3.106) and (3.107), respectively.

Furthermore, as mentioned in Section 4.2.2.1, for the bandwidth parameter involved in the kernel weights, we choose the recommended optimal value under the independent setup, namely $b = c_0 K^{-1/5}$, where $K = 100,\ 300$, with $c_0 = \sigma_z$.

Next we choose the regression and dynamic dependence parameters as

$$\boldsymbol{\beta} = (\beta_1, \beta_2)^\top = (0.5, 0.5)^\top; \text{and } \theta \equiv -3.0, -1.0, 1.0.$$

The data generation and estimation were done based on 1000 simulations.

The SCQL estimates for the function $\psi(\cdot)$ along with the true function are displayed in Figure 4.4 for the case with $K = 100$. The estimated functions appear to almost overlap the true function indicating good fitting. The SML estimates of $\boldsymbol{\beta}$ and

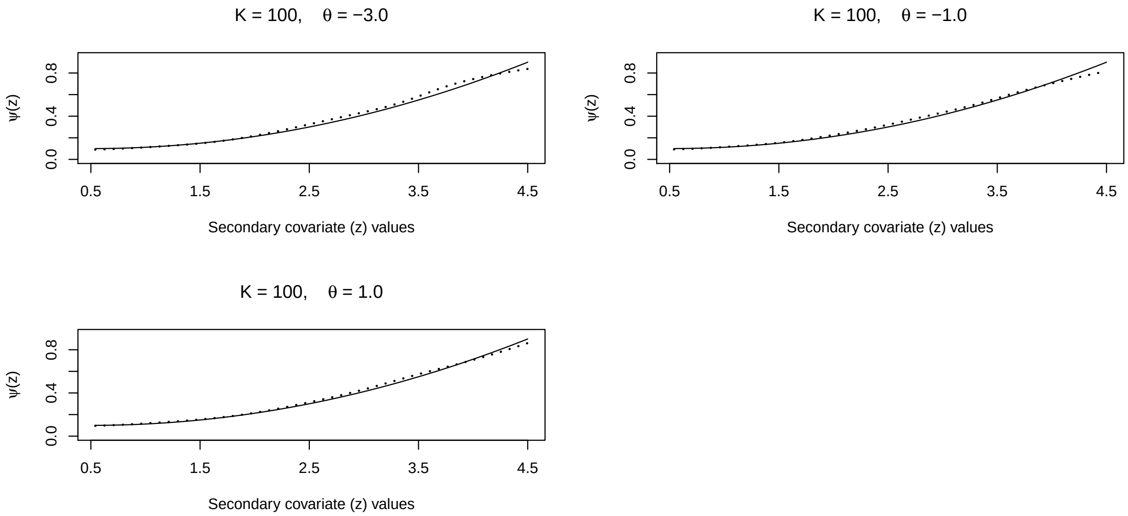

Figure 4.4: The plots for the true (thick curve) and estimated (dotted curve) nonparametric function $\psi(\cdot)$ for the SBDL model based on $\boldsymbol{\phi}$ estimate produced by the SML approach. The bandwidth $b = c_0 K^{-1/5}$ with $c_0 = \sigma_z \approx \left(\frac{z \text{ range}}{4}\right)$.

θ are shown in Table 4.5 for selected values of the dynamic dependence parameter $\theta \equiv -3.0, -1.0,$ and 1.0. The estimates are in general good agreement with the corresponding true values of the parameter. To be specific, the SML approach appears to produce almost unbiased estimates for the dynamic dependence parameter. For example, $\theta = -1.0$ is estimated as -1.04 and $\theta = 1.0$ is estimated as 1.00, when $K = 100$. As far as the estimation of regression effects is concerned, $\boldsymbol{\beta}$ estimates are less biased when the dynamic dependence is negative. For positive $\theta = 1.0$, the $\boldsymbol{\beta}$ estimates show some bias, but the bias gets smaller when K is increased. For example, for $\theta = 1.0$, the $\beta_2 = 0.5$ estimate is 0.55 when $K = 100$, but the estimate is found to be 0.51 when $K = 300$. Now because K is usually large in a longitudinal study, the proposed estimation approaches appear to be adequate in fitting the SBDL model (4.39)$-$(4.40).

K	Method	Quantity	β_1	β_2	θ
100			0.5	0.5	-3.0
	SML	SM	0.5272	0.5101	-3.1185
		SSE	0.3480	0.8523	0.3790
		MSE	0.1217	0.7258	0.1576
100			0.5	0.5	-1.0
	SML	SM	0.5302	0.5348	-1.0386
		SSE	0.3174	0.7340	0.2447
		MSE	0.1016	0.5395	0.0613
100			0.5	0.5	1.0
	SML	SM	0.5371	0.5547	0.9973
		SSE	0.3652	0.8954	0.3061
		MSE	0.1346	0.8039	0.0936
300			0.5	0.5	-3.0
	SML	SM	0.5094	0.5129	-3.04104
		SSE	0.2127	0.4933	0.2054
		MSE	0.0453	0.2432	0.0438
300			0.5	0.5	-1.0
	SML	SM	0.5075	0.5213	-1.0149
		SSE	0.1977	0.4344	0.1385
		MSE	0.0391	0.1890	0.0194
300			0.5	0.5	1.0
	SML	SM	0.5123	0.4968	0.9940
		SSE	0.2112	0.5063	0.1737
		MSE	0.0447	0.2561	0.0302

Table 4.5: Simulated means (SMs), simulated standard errors (SSEs) and mean square errors (MSEs) of the semi-parametric maximum likelihood (SML) estimates for the regression parameter $\boldsymbol{\beta}$ and dynamic dependence parameter θ, under the semi-parametric BDL model for selected parameter values with $K = 100, 300$, $n_i = 4$, and 1000 simulations. The bandwidth $b = c_0 K^{-1/5}$ with $c_0 = \sigma_z \approx \left(\frac{z \text{ range}}{4}\right)$.

4.2.5 An illustration: Fitting the SBDL model to the longitudinal infectious disease data

To illustrate the proposed semi-parametric BDL model (4.39)−(4.40), in this section, we reanalyze the respiratory infection (0 = no, 1 = yes) data earlier studied by some authors such as Zeger and Karim (1991), Diggle et al. (1994), Lin and Carroll (2001).

These binary data for the presence of respiratory infection were collected from 275 preschool-age children examined every quarter for up to six consecutive quarters. In our notation, y_{ij} indicates the infection status of the ith ($i = 1, \ldots, 275$) child collected on jth ($j = 1, \ldots, n_i$) quarter with max $n_i = 6$. A variety of primary covariates, namely vitamin A deficiency, sex, height, and stunting status, and a secondary covariate, namely the age of the child in the unit of month, were recorded. In our notation these primary and secondary covariates are denoted by $\boldsymbol{x}_{ij}(j)$ and z_{ij} respectively. Similar to the aforementioned studies, it is of main interest to find the effects ($\boldsymbol{\beta}$) of the primary covariates while fitting the secondary covariates through a smooth function $\psi(z_{ij})$. For the purpose, Lin and Carroll (2001, Section 8) for example, fitted a semi-parametric marginal model with binary means

$$\mu_{ij}(\boldsymbol{\beta}, \boldsymbol{x}_{ij}, \psi(z_{ij})) = \frac{\exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij}))}{1 + \exp(\boldsymbol{x}_{ij}^{\top}(t_{ij})\boldsymbol{\beta} + \psi(z_{ij}))},$$

with no specified correlation structures. These authors advocate for the use of the 'working' UNS (unstructured) correlations based GEE (generalized estimating equation) approach for efficient estimation of the function $\psi(\cdot)$ and other parameters as well. However, as we demonstrated in Section 4.1.4.1 in the context of SLDCP model that the UNS based GEE approach encounters efficiency drawbacks. For example, it was shown that SGEE(UNS) produces less efficient regression estimates than SGEE(I). Moreover, the GEE approach is not applicable to the present SBDL model (4.39)$-$(4.40) because unlike GEE models it is not a marginal model and the marginal means under this SBDL model contains the dynamic dependence parameter (see Eqn. (4.43)). Consequently, we do not include the 'working' correlations model based GEE approach to analyze this data set. Nevertheless, on top of fitting the proposed SBDL model (4.39)$-$(4.40), we also include the fitting of the SLDCP (semi-parametric LDCP) model as presented in Section 4.1.5, where the parameters including the correlation index ρ, and the nonparametric function were estimated using a SGQL (semi-parametric GQL) approach. These SGQL estimates of the parameters along with the SML estimates for the proposed (main) SBDL model (4.40) are displayed in Table 4.6. Note that the simulation study in Section 4.2.4 showed

	Models			
	SLDCP		SBDL	
Primary covariates	SGQL Estimate	SE	SML Estimate	SE
Vitamin A deficiency	0.576	0.448	0.567	0.522
Seasonal Cosine	-0.579	0.170	-0.772	0.224
Seasonal Sine	-0.156	0.168	-0.174	0.179
Sex	-0.515	0.227	-0.467	0.274
Height	-0.027	0.025	0.004	0.029
Stunting	0.464	0.407	0.697	0.493
Dynamic dependence parameter ρ	0.020	–		
Dynamic dependence parameter θ	–	–	0.261	0.381

Table 4.6: Primary regression effect estimates (SML) along with their standard errors for the respiratory infectious data under the semi-parametric BDL (SBDL) model (4.40). The SGQL estimates under a SLDCP model are also given. The bandwidth is used as $b = \left(\frac{\text{age range}}{4}\right) K^{-1/5}$, where K is the number of individuals in the relevant data.

that the SML and SCQL approaches work very well in estimating the nonparametric function $\psi(\cdot)$ and the parameters $\boldsymbol{\beta}$ and θ involved in the SBDL model (4.40). The SCQL estimate for the nonparametric function of the SBDL model for fitting the infectious disease data is displayed in Figure 4.5. Unlike in Lin and Carroll (2001) these estimates in general show a linear negative effect of age rather than any quadratic effect.

Further, we remark that because in a longitudinal study, the mean (and variance as well) function of the data usually change with regard to time mainly due to the influence of time dependent primary and secondary covariates, it may not be enough

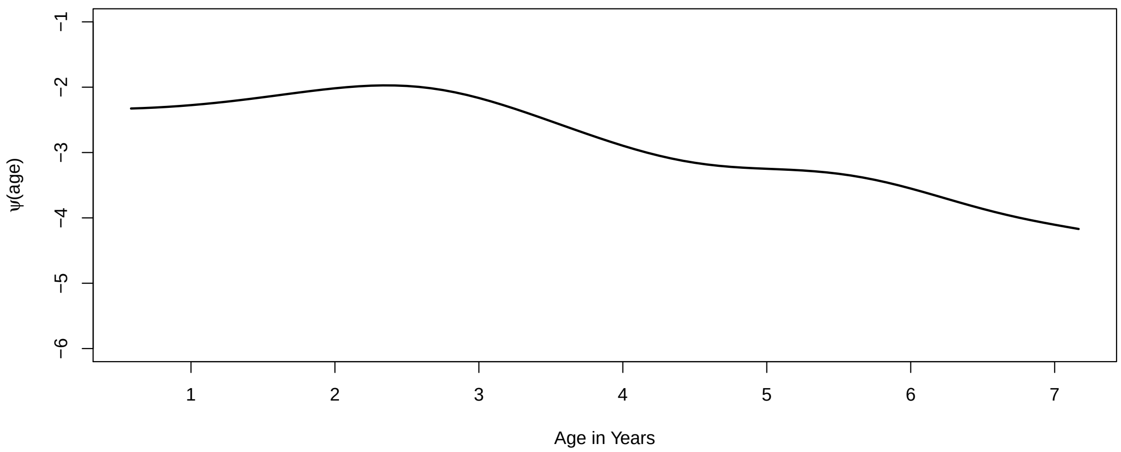

Figure 4.5: Estimated $\psi(\cdot)$ function for the unbalanced infectious disease data using the semi-parametric BDL (SBDL) model. The bandwidth $b = \left(\frac{\text{age range}}{4}\right) K^{-1/5}$, where K is the number of individuals in the relevant data.

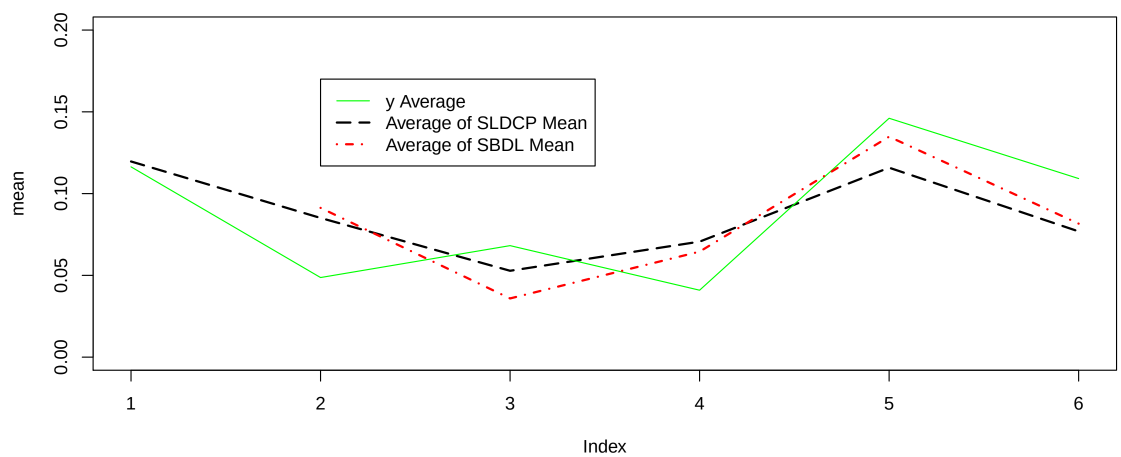

Figure 4.6: The average of the estimated means under the SBDL (4.40) and SLDCP (4.4) models, and the average of y values at each longitudinal index (time) point for the unbalanced infectious disease data. The bandwidth $b = \left(\frac{\text{age range}}{4}\right) K^{-1/5}$, where K is the number of individuals in the relevant data.

to only examine the effects of primary covariates in such a study. For this reason, we have computed the averages of the binary data along with their estimated means over the time range under both SLDCP and SBDL models. For a given time j, these averages are

$$\begin{aligned}
\bar{y}_j &= \frac{\sum_{i=1}^{K} y_{ij}}{K}, \\
\hat{\mu}_j(\text{for SLDCP model}) &= \frac{\sum_{i=1}^{K} \hat{\mu}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\rho}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}, z_{ij}))}{K}, \text{ and} \\
\hat{\pi}_j(\text{for SBDL model}) &= \frac{\sum_{i=1}^{K} \hat{\pi}_{ij}(\hat{\boldsymbol{\beta}}, \hat{\theta}, \boldsymbol{x}_{ij}, \hat{\psi}(\hat{\boldsymbol{\beta}}, \hat{\theta}, z_{ij}))}{K},
\end{aligned}$$

respectively. The display of these averages in Figure 4.6 shows that the fitted means under the SBDL model (in dotted red) are closer to the mean functions of the binary observations (in solid green) than the fitted means under the SLDCP model, except for the time around 3rd quarter. However, this SBDL model does not produce marginal means used by Lin and Carroll (2001) for example, to interpret the regression effects. As shown in Table 4.6, except for the effect of Vitamin A deficiency, the estimated effect values for the remaining primary covariates are generally different under SLDCP and SBDL models. For example, stunting covariate affect the presence of infection with coefficient 0.46 under the SLDCP model but with 0.69 under the SBDL model. Because the estimated nonparametric functions show a negative linear effect of age on the responses, we have also fitted a parametric LDCP (PLDCP) model by treating age as an additional primary covariate (results are shown in Table 4.4). However, the effects of some of the covariates such as sex and stunting are quite different under this PLDCP model as compared to the semi-parametric LDCP (SLDCP) model.

In summary, because SBDL model appears to fit the mean function of the observations over time better than the SLDCP model, we chose to interpret the effects of the primary covariates under the SBDL model. To be specific, the Vitamin A deficiency (yes/no) has a large positive effect 0.57 on the probability of having respiratory infection in a child. The negative value -0.47 for the sex effect shows that female child (coded as 1) has smaller probability of having respiratory infection. As far as the nonparametric function effect is concerned, the estimated function under the SBDL model shows that as age increases the infection probability gets decreased.

Chapter 5

Semi-parametric dynamic mixed models for longitudinal binary data

In the previous chapter we have introduced two semi-parametric dynamic fixed models for binary data, namely the SLDCP (4.3)−(4.4) and SBDL (4.39)−(4.40) models. However, as discussed in Chapter 3 for count data, there may be situations where suitable random effects involved in semi-parametric linear predictor may explain the data better than the fixed model. In this chapter, we consider the mixed model extension of the binary fixed models discussed in Chapter 4. However, because the SBDL model, as opposed to the SLDCP model, produces mean functions based on the past history which is more practical, in this chapter we generalize the SBDL model only. We refer to such model as the SBDML (semi-parametric binary dynamic mixed logit) model. One could also extend the SLDCP model to the SLDMCP (semi-parametric linear dynamic mixed conditional probability) model, but we did not enclose this generalization to save space. As another reason for our preference of the SBDL model to SLDCP model, the SBDL model allows unrestricted dynamic dependence parameter values, whereas some special care is needed about the restriction of the dynamic dependence parameters under the SLDCP model.

We now generalize the SBDL model ((4.39)−(4.40)) to the mixed model setup as follows. Similar to the mixed model for count data (3.3), we simply add a random effect $\tau_i^* = \sigma_\tau \tau_i$ with $\tau_i \overset{i.i.d.}{\sim} N(0,1)$ to the SBDL model as

$$\Pr(y_{ij}=1|y_{i,j-1},\mathbf{x}_{ij},\tau_i) = \begin{cases} \frac{\exp(\mathbf{x}_{i1}^\top(t_{i1})\boldsymbol{\beta}+\psi(z_{i1})+\sigma_\tau\tau_i)}{1+\exp(\mathbf{x}_{i1}^\top(t_{i1})\boldsymbol{\beta}+\psi(z_{i1})+\sigma_\tau\tau_i)} = p_{i10}^*, & \text{for } j=1 \\ \frac{\exp(\mathbf{x}_{ij}^\top(t_{ij})\boldsymbol{\beta}+\theta y_{i,j-1}+\psi(z_{ij})+\sigma_\tau\tau_i)}{1+\exp(\mathbf{x}_{ij}^\top(t_{ij})\boldsymbol{\beta}+\theta y_{i,j-1}+\psi(z_{ij})+\sigma_\tau\tau_i)} = p_{ijy_{i,j-1}}^*, & \text{for } j=2,\cdots,n_i\,. \end{cases} \tag{5.1}$$

Note that this SBDML model (5.1) reduces to the BDML model (Sutradhar, 2011, Chapter 9), namely

$$\Pr(y_{ij}=1|y_{i,j-1},\mathbf{x}_{ij},\tau_i) = \begin{cases} \frac{\exp(\mathbf{x}_{i1}^\top(t_{i1})\boldsymbol{\beta}+\sigma_\tau\tau_i)}{1+\exp(\mathbf{x}_{i1}^\top(t_{i1})\boldsymbol{\beta}+\sigma_\tau\tau_i)}, & \text{for } j=1 \\ \frac{\exp(\mathbf{x}_{ij}^\top(t_{ij})\boldsymbol{\beta}+\theta y_{i,j-1}+\sigma_\tau\tau_i)}{1+\exp(\mathbf{x}_{ij}^\top(t_{ij})\boldsymbol{\beta}+\theta y_{i,j-1}+\sigma_\tau\tau_i)}, & \text{for } j=2,\cdots,n_i\,, \end{cases} \tag{5.2}$$

when the nonparametric function $\psi(\cdot)$ is ignored. This BDML model has been studied by many authors such as Heckman (1979), Manski (1987), Honoré and Kyriazidou (2000). For a detailed study including the estimation of the parameters of this BDML model, we refer to Sutradhar (2011, Chapter 9).

The SBDML model (5.1) is quite general. The longitudinal correlations among repeated responses arise due to the dynamic dependence of a current response (y_{ij}) on the past response ($y_{i,j-1}$), and each response (y_{ij}) is also influenced by a latent effect (τ_i) of the ith individual causing overdispersion and hence structural correlations. This model (5.1) is quite different than some existing mixed effect models (Breslow and Clayton, 1993, Lin and Carroll, 2006), where longitudinal correlations are assumed to be generated only through random effects. In (5.1), the dynamic dependence conditional on the random effects introduces longitudinal correlations among the repeated binary responses. For this reason, the random effects model considered by Lin and Carroll (2006, Example 3) would produce correlations without lag dependence, hence it should not be used as longitudinal correlations. For example, under their model, the $\mathrm{Cov}(Y_{ij}, Y_{ik})$ does not depend on lag $|k-j|$, instead they remain similar or the same for any small or large value of $|k-j|$, which is generally inappropriate for repeated responses. In contrast, the introduction of dynamic dependence parameter θ in model (5.1) results in a more flexible lag-dependent correlation structure, which may also contain the influence from the random effects through σ_τ^2.

5.1 Basic properties of the proposed SBDML model (5.1)

As in other chapters of the thesis, the primary covariates $\mathbf{x}_{ij}$ are always considered fixed. So for simplicity, we will drop the conditioning on $\mathbf{x}_{ij}$. Notice that the SBDML model (5.1) can also be written in the form of

$$\mathrm{E}[Y_{ij}|y_{i,j-1}, \tau_i] = p^*_{ij1} y_{i,j-1} + p^*_{ij0}(1 - y_{i,j-1})\,, \tag{5.3}$$

where $p^*_{ij0} = \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \psi(z_{ij}) + \sigma_\tau \tau_i)/[1 + \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \psi(z_{ij}) + \sigma_\tau \tau_i)]$ and $p^*_{ij1} = \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \theta + \psi(z_{ij}) + \sigma_\tau \tau_i)/[1 + \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \theta + \psi(z_{ij}) + \sigma_\tau \tau_i)]$. It follows that the mean of Y_{ij} conditional on τ_i, which is denoted by $\mu^*_{ij}(\tau_i) = \mathrm{E}[Y_{ij}|\tau_i], j = 1, \cdots, n_i$, can be written as

$$\mu^*_{ij}(\tau_i) = \begin{cases} p^*_{i10}, & \text{for } j = 1 \\ p^*_{ij0} + \left(p^*_{ij1} - p^*_{ij0}\right) \mu^*_{i,j-1}(\tau_i), & \text{for } j = 2, \cdots, n_i \,, \end{cases} \tag{5.4}$$

yielding the unconditional mean $\mu_{ij} = \mathrm{E}(Y_{ij})$ and variance $\sigma_{ijj} = \mathrm{Var}(Y_{ij})$ as

$$\mu_{ij} = \int_{-\infty}^{\infty} \mu^*_{ij}(\tau_i)\phi(\tau_i)d\tau_i \simeq \frac{1}{N}\sum_{w=1}^{N} \mu^*_{ij}(\tau_{iw}) \quad \text{and} \quad \sigma_{ijj} = \mu_{ij}\,(1-\mu_{ij}) \tag{5.5}$$

respectively, where $\phi(\cdot)$ is the standard normal density, N is a large number such as $N = 1000$, and $\tau_{iw}, w = 1, \cdots, N$, is a random sample from a standard normal distribution. Notice that the recursive nature of formula (5.4) implies that the unconditional mean μ_{ij} depends not only on the present covariate values $(\mathbf{x}_{ij}, z_{ij})$, but also on the past covariate values from $(\mathbf{x}_{i,j-1}, z_{i,j-1})$ to $(\mathbf{x}_{i1}, z_{i1})$. This shows a major difference between the SBDML model (5.1) and the semi-parametric fixed models studied by other authors such as Severini and Staniswalis (1994), Lin and Carroll (2001).

As far as the second order moments are concerned, one may first write

$$\mathrm{E}[(Y_{ij} - \mu^*_{ij})(Y_{ik} - \mu^*_{ik})|\tau_i] = \mu^*_{ij}(1 - \mu^*_{ij}) \prod_{l=j+1}^{k} (p^*_{il1} - p^*_{il0}),$$

yielding

$$\lambda^*_{ijk}(\tau_i) = \mathrm{E}[Y_{ij}Y_{ik}|\tau_i] \;=\; \mu^*_{ij}(\tau_i)\left(1 - \mu^*_{ij}(\tau_i)\right) \prod_{l=j+1}^{k} (p^*_{il1} - p^*_{il0}) + \mu^*_{ij}(\tau_i)\mu^*_{ik}(\tau_i)\,. \tag{5.6}$$

It then follows that the unconditional second order moments are given by

$$\lambda_{ijk} \;\equiv\; \mathrm{E}[Y_{ij}Y_{ik}] \;=\; \int_{-\infty}^{\infty} \lambda^*_{ijk}(\tau_i)\phi(\tau_i)d\tau_i \;\simeq\; \frac{1}{N}\sum_{w=1}^{N} \lambda^*_{ijk}(\tau_{iw})\,, \tag{5.7}$$

for any $j \neq k,\ j,k = 1,\cdots,n_i$; yielding $\text{Cov}(Y_{ij}, Y_{ik}) = \lambda_{ijk} - \mu_{ij}\mu_{ik}$, and hence correlations as

$$\text{Corr}(Y_{ij}, Y_{ik}) \equiv \frac{\lambda_{ijk} - \mu_{ij}\mu_{ik}}{\sqrt{\sigma_{ijj}\sigma_{ikk}}}. \tag{5.8}$$

Furthermore, when needed, by using

$$\text{E}[Y_{ij} - \mu^*_{ij}|y_{i,j-1}, \tau_i] = \left(p^*_{ij1} - p^*_{ij0}\right)\left(y_{i,j-1} - \mu^*_{i,j-1}\right) \tag{5.9}$$

for $j \geq 2$, one can derive other higher order moments. These moments will be necessary for the construction of SGQL estimating equations for the parameters of the model (see Section 5.2).

Note that in practice, it is necessary to understand the mean, variance and correlation of the responses. For this reason, one must estimate the means and variances by (5.5) and correlations by (5.8) consistently, which requires the consistent estimation of the parameters, namely, $\boldsymbol{\beta}$, θ and σ^2_τ, and $\psi(\cdot)$. This estimation is discussed in details in Section 5.2. The asymptotic properties of the estimators are given in Section 5.3.

5.2 Estimation

The estimation of the parameters of the BDML (binary dynamic mixed logit) model has been carried out by Sutradhar et al. (2010) using both GQL and ML approaches, GQL being the simpler but competitive. The proposed SBDML model (5.1) is a generalization of the BDML model (5.2) to the semi-parametric setup. Note that because the parameters of this SBDML model cannot be estimated consistently without estimating the function $\psi(\cdot)$ consistently, this makes the whole estimation procedure more complex. In a mixed model framework, this has not been addressed in the literature so far. In this section, for the estimation of the parameters, we continue to explore the GQL and ML approaches following Sutradhar et al. (2010). To be specific, we modify their estimation approaches to accommodate the fact that the estimate of the nonparametric function is obtained for known values of the parameters.

We remark that some authors (Severini and Staniswalis, 1994, Lin and Carroll, 2001) dealt with estimation of the SBL (semi-parametric binary logit) marginal model as opposed to the recursive semi-parametric model (4.39)−(4.40). These authors used the SGEE (semi-parametric generalized estimating equation) approach for marginal models, which is not applicable for the present situation because the mean and variance under the present model (5.1) include the dynamic dependence parameters. Moreover, these authors did not consider any mixed effects in their marginal models.

5.2.1 Estimation of the nonparametric function $\psi(\cdot)$: A SCQL approach

In marginal model setup, the nonparametric function involved in the semi-parametric models has been estimated by some authors (Severini and Staniswalis, 1994, Lin and Carroll, 2001) using WGEE (weighted generalized estimating equation) approach where estimating equation was constructed by assuming a "working" correlation matrix. However, in the context of fitting the generalized linear longitudinal fixed model to count data, Sutradhar et al. (2016) [see also Chapter 2] have demonstrated that a "working" independence assumption-based GEE, i.e., GEE(I) approach, still produces a consistent estimate for the nonparametric function, and this approach is much simpler.

Note that as opposed to the marginal models, we exploited the aforementioned idea of applying independence assumption for consistent estimation of nonparametric function in Chapter 4 under the dynamic fixed model setup for binary data. More specifically, under the SBDL (semi-parametric binary dynamic logit) model we used SCQL (semi-parametric conditional QL) approach which makes the present response independent to past responses conditionally. In this section, we follow the SCQL approach but accommodate the random effect variance involved in the present SBDML model. Thus we use all y_{ij} conditional on $y_{i,j-1}$ $(j = 2, \cdots, n_i)$ to construct the desired SCQL estimating equation. For $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma_\tau^2)^\top$ we first compute the marginal mean at time point $j = 1$ and the conditional means for $j = 2, \cdots, n_i$, as

$$\begin{aligned}
\Pr(y_{i1} = 1) = \mathrm{E}[Y_{i1}] &= \int_{-\infty}^{\infty} p^*_{i10}(\boldsymbol{\alpha}, \psi(z_{i1}), \tau_i)\phi(\tau_i)d\tau_i = p^\dagger_{i10}(\boldsymbol{\alpha}, \psi(z_{i1})) \\
\Pr(y_{ij} = 1|y_{i,j-1}) = \mathrm{E}[Y_{ij}|y_{i,j-1}] &= \int_{-\infty}^{\infty} p^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\phi(\tau_i)d\tau_i, \quad j = 2, \cdots, n_i \\
&= p^\dagger_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij})), \quad \text{say}, \qquad (5.10)
\end{aligned}$$

where $p^*_{ijy_{i,j-1}}(\cdot)$ is given in (5.1), and $\phi(\cdot)$ is the probability density function (pdf) of the standard normal distribution. We remark that the conditional probability $p^\dagger_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}))$ is different than the conditional probability $p_{i,j|j-1}(\boldsymbol{\beta}, \theta, \psi(z_{ij}))$ used under the fixed model. It then follows that the variance at $j = 1$ and the variances at $j = 2, \cdots, n_i$ conditional on the previous responses have the forms

$$\begin{aligned}
\mathrm{var}(Y_{i1}) &= p^\dagger_{i10}(\boldsymbol{\alpha}, \psi(z_{i1}))\left[1 - p^\dagger_{i10}(\boldsymbol{\alpha}, \psi(z_{i1}))\right] \\
\mathrm{var}(Y_{ij}|y_{i,j-1}) &= p^\dagger_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}))\left[1 - p^\dagger_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}))\right], \quad j = 2, \cdots, n_i \\
&= \sigma_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij})), \quad \text{say}. \qquad (5.11)
\end{aligned}$$

Note that for notational convenience, we may define a dummy response y_{i0} and set $y_{i0} = 0$ for all $i = 1, \cdots, K$. Then the SCQL estimating equation for estimating the nonparametric function at covariate value z is given by

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\, v_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z))\, \sigma^{-1}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z)) \left[y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z))\right]$$
$$= f(\psi(z),\boldsymbol{\alpha}) = 0\,, \tag{5.12}$$

where

$$\begin{aligned} v_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z)) &= \frac{\partial p^{\dagger}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z))}{\partial \psi(z)} \\ &= \int_{-\infty}^{\infty} p^{*}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z),\tau_i)\left[1 - p^{*}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z),\tau_i)\right]\phi(\tau_i)d\tau_i\,, \end{aligned} \tag{5.13}$$

and $w_{ij}(z)$ is referred to as the so-called kernel weight defined as

$$w_{ij}(z) = h_{ij}(\frac{z - z_{ij}}{b})\Big/\sum_{l=1}^{K}\sum_{u=1}^{n_l} h_{lu}(\frac{z - z_{lu}}{b})\,. \tag{5.14}$$

Here, to avoid confusing with other notations in this chapter, we use $h_{ij}(\cdot)$ to denote the kernel density $p_{ij}(\cdot)$ defined in (3.12)−(3.14), with b as a suitable bandwidth parameter.

For known $\boldsymbol{\alpha}$, one may then solve the estimating equation (5.12) by using the iterative equation given by

$$\hat{\psi}(z,\boldsymbol{\alpha})_{(r+1)} = \hat{\psi}(z,\boldsymbol{\alpha})_{(r)} - \left[\left\{f'_{\psi(z)}(\psi(z),\boldsymbol{\alpha})\right\}^{-1} f(\psi(z),\boldsymbol{\alpha})\right]_{|\psi(z)=\hat{\psi}(z,\boldsymbol{\alpha})_{(r)}}\,, \tag{5.15}$$

where (r) indicates the rth iteration and $f'_{\psi(z)}(\psi(z),\boldsymbol{\alpha})$ has the formula

$$f'_{\psi(z)}(\psi(z),\boldsymbol{\alpha}) \;\simeq\; -\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\, v^{2}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z))\, \sigma^{-1}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z))\,.$$

For notational simplicity, $p^{\dagger}_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z))$ and $v_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z))$ will be abbreviated as $p^{\dagger}_{ijy_{i,j-1}}(z)$ and $v_{ijy_{i,j-1}}(z)$, respectively.

We remark that conditional on τ_i's, the conditional SCQL [see (4.48)] estimating equation is written as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\frac{\partial p^{*}_{ijy_{i,j-1}}(z)}{\partial \psi(z)}\frac{y_{ij} - p^{*}_{ijy_{i,j-1}}(z)}{p^{*}_{ijy_{i,j-1}}(z)\left(1 - p^{*}_{ijy_{i,j-1}}(z)\right)} = 0$$
$$\Rightarrow \sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\left(y_{ij} - p^{*}_{ijy_{i,j-1}}(z)\right) = 0\,.$$

Then by taking expectation over τ_i's, we obtain the semi-parametric conditional moment estimating equation for $\psi(z)$ as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}\int w_{ij}(z)\left(y_{ij} - p^{*}_{ijy_{i,j-1}}(z)\right)\phi(\tau_i)d\tau_i = \sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\left(y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z)\right) = 0\,. \tag{5.16}$$

Simulation study shows that (5.12) gives better estimating results than (5.16) does. So in later part of this chapter, we concentrate on the estimating approaches with $\psi(z)$ estimated by (5.12).

5.2.2 Joint estimation of the regression, dynamic dependence and over-dispersion index parameters

In Section 5.2.1, the SCQL estimate $\hat{\psi}(z, \boldsymbol{\alpha})$ for the function $\psi(z)$ was obtained for known $\boldsymbol{\alpha}$. In following sections, we demonstrate how $\boldsymbol{\alpha}$ can be consistently estimated by using the aforementioned SGQL and SML approaches.

5.2.2.1 A SGQL estimation approach

Because $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma_\tau^2)^\top$, we construct a first and second order (pairwise products) response-based quasi-likelihood estimating equation (Sutradhar et al., 2010) for its estimation. Let $\mathbf{u}_i = (\mathbf{y}_i^\top, \mathbf{s}_i^\top)^\top$ represent this vector with $\mathbf{y}_i = (y_{i1}, \cdots, y_{in_i})^\top$ as the n_i-dimensional vector of responses for the ith individual and $\mathbf{s}_i = (y_{i1}y_{i2}, \cdots, y_{ij}y_{ik}, \cdots, y_{i,n_i-1}y_{in_i})^\top$ be the $(n_i-1)n_i/2$-dimensional vector of distinct pairwise products of the n_i responses. Let $\boldsymbol{\lambda}_i = \mathrm{E}[\mathbf{U}_i] = (\mathrm{E}[\mathbf{Y}_i^\top], \mathrm{E}[\mathbf{S}_i^\top])^\top$ be the expectation of the vector $\mathbf{u}_i$, which is already computed in Section 5.1. To be specific, $\mu_{ij} = \mathrm{E}[Y_{ij}]$ and $\lambda_{ijk} = \mathrm{E}[Y_{ij}Y_{ik}]$ are known by (5.5) and (5.7), respectively. Next, let $\boldsymbol{\Omega}_i$ be the $\{n_i(n_i+1)/2 \times n_i(n_i+1)/2\}$ covariance matrix of $\mathbf{u}_i$ for the ith individual. In the SGQL approach, one essentially minimizes the so-called generalized squared distance

$$\sum_{i=1}^{K} (\mathbf{u}_i - \boldsymbol{\lambda}_i)^\top \boldsymbol{\Omega}_i^{-1} (\mathbf{u}_i - \boldsymbol{\lambda}_i) \tag{5.17}$$

to estimate the desired parameters of the model. This provides the estimating equation (5.18) below for $\boldsymbol{\alpha}$ after some modifications.

For the present semi-parametric model, we note that the true function $\psi(\cdot)$ is unknown, instead its estimator $\hat{\psi}(\cdot, \boldsymbol{\alpha})$ is used, where $\hat{\psi}(\cdot, \boldsymbol{\alpha})$ was obtained by solving the SCQL estimating equation (5.12) for known $\boldsymbol{\alpha}$. However, in practice $\boldsymbol{\alpha}$ are unknown parameters. Note that the means and second order moments under model (5.1) are defined as $\mu_{ij}(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))$ $(j = 1, \cdots, n_i)$ and $\lambda_{ijk}(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))$ $(j \neq k; j, k = 1, \cdots, n_i)$, respectively, with $\mathbf{z}_i = (z_{i1}, \cdots, z_{in_i})^\top$, but the vector function $\psi(\mathbf{z}_i)$ is estimated as $\hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})$. For this reason, we modify the notations for the moments when $\psi(\mathbf{z}_i)$ is replaced with $\hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})$ as $\tilde{\mu}_{ij}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))$ and $\tilde{\lambda}_{ijk}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))$, that is, we add a symbol, say tilde for the quantities containing the estimated nonparametric function $\hat{\psi}(\cdot, \boldsymbol{\alpha})$, and write out the parameter and nonparametric function dependence explicitly.

Minimization of the generalized squared distance (5.17) for the estimation of $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma_\tau^2)^\top$ leads to the SGQL estimating equations for $\boldsymbol{\alpha}$ as

$$\sum_{i=1}^{K} \frac{\partial \tilde{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}} \tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})) \left[\mathbf{u}_i - \tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))\right] \;=\; 0\,, \qquad (5.18)$$

which may be solved iteratively by using

$$\begin{aligned}\hat{\boldsymbol{\alpha}}(r+1) = \hat{\boldsymbol{\alpha}}(r) + &\left[\sum_{i=1}^{K} \frac{\partial \tilde{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}} \tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})) \frac{\partial \tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}^\top}\right]_r^{-1} \times \\ &\left[\sum_{i=1}^{K} \frac{\partial \tilde{\boldsymbol{\lambda}}_i^\top(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}} \tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})) \left[\mathbf{u}_i - \tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))\right]\right]_r ,\end{aligned} \qquad (5.19)$$

where $[\cdot]_r$ denotes that the quantity in the parenthesis is evaluated at $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}(r)$, the value of $\boldsymbol{\alpha}$ obtained from the rth iteration. Let $\hat{\boldsymbol{\alpha}}_{SGQL}$ denote the solution of (5.18) obtained by (5.19).

Notice that unlike in the longitudinal mixed model case (Sutradhar et al., 2010), the computation for the gradient functions in the present semi-parametric longitudinal mixed model setup requires special care for additional derivatives of the estimated nonparametric function with respect to $\boldsymbol{\alpha}$. For convenience, we give below the main formulas for the gradients with details for the additional derivatives in the Appendix A.3. To be specific, for $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma_\tau^2)^\top \equiv (\boldsymbol{\alpha}_1^\top, \alpha_2, \alpha_3)^\top$ and a large integer N,

$$\frac{\partial \tilde{\mu}_{ij}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}} \;=\; \frac{1}{N} \sum_{w=1}^{N} \begin{pmatrix} \partial/\partial \boldsymbol{\alpha}_1 \\ \partial/\partial \alpha_2 \\ \partial/\partial \alpha_3 \end{pmatrix} \tilde{\mu}_{ij}^*(\tau_{iw}, \boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})) \qquad (5.20)$$

by (5.5), and

$$\frac{\partial \tilde{\lambda}_{ijk}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \boldsymbol{\alpha}} = \frac{1}{N} \sum_{w=1}^{N} \begin{pmatrix} \partial / \partial \boldsymbol{\alpha}_1 \\ \partial / \partial \alpha_2 \\ \partial / \partial \alpha_3 \end{pmatrix} \tilde{\lambda}^*_{ijk}(\tau_{iw}, \boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})) \tag{5.21}$$

by (5.7). In (5.20),

$$\frac{\partial \tilde{\mu}^*_{ij}(\tau_{iw}, \boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))}{\partial \alpha_m} = \begin{cases} c_{m,1} & \text{if } j = 1 \\ c_{m,j} + d_{m,j} \tilde{\mu}^*_{i,j-1} + \left(\tilde{p}^*_{ij1} - \tilde{p}^*_{ij0}\right) \frac{\partial \tilde{\mu}^*_{i,j-1}}{\partial \alpha_m} & \text{if } 2 \leq j \leq n_i \, , \end{cases} \tag{5.22}$$

with

$$c_{m,j} = \begin{cases} \tilde{p}^*_{ij0} \left(1 - \tilde{p}^*_{ij0}\right) \left[\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_m}\right] & \text{for } m = 1 \\ \tilde{p}^*_{ij0} \left(1 - \tilde{p}^*_{ij0}\right) \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \alpha_m} & \text{for } m = 2 \\ \tilde{p}^*_{ij0} \left(1 - \tilde{p}^*_{ij0}\right) \left[\frac{1}{2\sigma_\tau} \tau_{iw} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \alpha_m}\right] & \text{for } m = 3 \, , \end{cases}$$

and

$$d_{m,j} = \begin{cases} \left[\tilde{p}^*_{ij1} \left(1 - \tilde{p}^*_{ij1}\right) - \tilde{p}^*_{ij0} \left(1 - \tilde{p}^*_{ij0}\right)\right] \left[\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}_m}\right] & \text{for } m = 1 \\ \tilde{p}^*_{ij1} \left(1 - \tilde{p}^*_{ij1}\right) \left[1 + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \alpha_m}\right] - \tilde{p}^*_{ij0} \left(1 - \tilde{p}^*_{ij0}\right) \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \alpha_m} & \text{for } m = 2 \\ \left[\tilde{p}^*_{ij1} \left(1 - \tilde{p}^*_{ij1}\right) - \tilde{p}^*_{ij0} \left(1 - \tilde{p}^*_{ij0}\right)\right] \left[\frac{1}{2\sigma_\tau} \tau_{iw} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \alpha_m}\right] & \text{for } m = 3 \, . \end{cases}$$

In (5.21), for $j < k$,

$$\begin{aligned} \frac{\partial \tilde{\lambda}^*_{ijk}}{\partial \alpha_m} &= \left(1 - 2\tilde{\mu}^*_{ij}\right) \frac{\partial \tilde{\mu}^*_{ij}}{\partial \alpha_m} \prod_{l=j+1}^{k} \left(\tilde{p}^*_{il1} - \tilde{p}^*_{il0}\right) + \left[\tilde{\mu}^*_{ij} \frac{\partial \tilde{\mu}^*_{ik}}{\partial \alpha_m} + \tilde{\mu}^*_{ik} \frac{\partial \tilde{\mu}^*_{ij}}{\partial \alpha_m}\right] \\ &\quad + \tilde{\mu}^*_{ij} \left(1 - \tilde{\mu}^*_{ij}\right) \sum_{u=j+1}^{k} d_{m,u} \prod_{\substack{l \neq u \\ l=j+1}}^{k} \left(\tilde{p}^*_{il1} - \tilde{p}^*_{il0}\right) . \end{aligned} \tag{5.23}$$

As indicated above, these main gradient functions in (5.20)-(5.23) contain the derivative of the estimated function $\hat{\psi}(z_{ij}, \boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \alpha_2, \alpha_3)^\top$. The formula for this additional derivative, i.e., $\partial\hat{\psi}(z_{ij}, \boldsymbol{\alpha})/\partial\boldsymbol{\alpha}$ is lengthy, and for convenience are given in the Appendix A.3.

Next, all moments needed to compute $\tilde{\boldsymbol{\Omega}}_i$ matrix in (5.18) are given in Section 5.1 and Appendix A.1, with $\psi(\cdot)$ replaced by its estimates $\hat{\psi}(\cdot, \boldsymbol{\alpha})$. For example, for $j < k$

$$\tilde{\mathrm{Cov}}(Y_{ij}, Y_{ik}) \;=\; \tilde{\lambda}_{ijk}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha})) - \tilde{\mu}_{ij}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))\, \tilde{\mu}_{ik}(\boldsymbol{\alpha}, \hat{\psi}(\mathbf{z}_i, \boldsymbol{\alpha}))\,.$$

5.2.2.2 The semi-parametric maximum likelihood (SML) method for $\boldsymbol{\alpha}$ estimation: An alternative estimation approach

In last section, we have developed a moments-based SGQL approach for the estimation of the parameters in $\boldsymbol{\alpha}$. It, however, appears from the model (5.1) that we can also use the well-known likelihood approach for this estimation, but, we caution that the computations will be much more involved as compared to the SGQL approach. Nevertheless, for the sake of completeness, we develop the likelihood estimating equations for the components of $\boldsymbol{\alpha}$ in order to examine the relative efficiency of the simpler SGQL approach.

To derive the likelihood estimating equations, we first construct the likelihood function for the semi-parametric mixed model (5.1) using $\hat{\psi}(z_{ij}, \boldsymbol{\alpha})$ for $\psi(z_{ij})$, as

$$\begin{aligned} \tilde{L}(\boldsymbol{\beta}, \theta, \sigma_\tau^2, \hat{\psi}(\cdot, \boldsymbol{\alpha})) = \prod_{i=1}^{K} \int_{-\infty}^{\infty} & \left[\frac{\exp([\mathbf{x}_{i1}^\top \boldsymbol{\beta} + \hat{\psi}(z_{i1}, \boldsymbol{\alpha}) + \sigma_\tau \tau_i] y_{i1})}{1 + \exp([\mathbf{x}_{i1}^\top \boldsymbol{\beta} + \hat{\psi}(z_{i1}, \boldsymbol{\alpha}) + \sigma_\tau \tau_i] y_{i1})} \right. \\ & \left. \times \prod_{j=2}^{n_i} \frac{\exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} y_{ij} + \theta y_{i,j-1} y_{ij} + \hat{\psi}(z_{ij}, \boldsymbol{\alpha}) y_{ij} + \sigma_\tau \tau_i y_{ij})}{1 + \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \theta y_{i,j-1} + \hat{\psi}(z_{ij}, \boldsymbol{\alpha}) + \sigma_\tau \tau_i)} \right] \phi(\tau_i)\, d\tau_i\,. \end{aligned} \quad (5.24)$$

Recall that we used $y_{i0} = 0$ as a conventional notation. One may then write the log-likelihood by (5.24) as

$$\log \tilde{L}(\boldsymbol{\beta}, \theta, \sigma_\tau^2, \hat{\psi}(\cdot, \boldsymbol{\alpha})) = \sum_{i=1}^{K} \left\{ \sum_{j=1}^{n_i} \left[\mathbf{x}_{ij}^\top \boldsymbol{\beta} y_{ij} + \theta y_{i,j-1} y_{ij} + \hat{\psi}(z_{ij}, \boldsymbol{\alpha}) y_{ij} \right] + \log J_i \right\}, \tag{5.25}$$

where $J_i = \int_{-\infty}^{\infty} \exp(\sigma_\tau \tau_i s_i) \Delta_i \phi(\tau_i) d\tau_i$ with $s_i = \sum_{j=1}^{n_i} y_{ij}$, and $\Delta_i = \prod_{j=1}^{n_i} \{1 + \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta} + \theta y_{i,j-1} + \hat{\psi}(z_{ij}, \boldsymbol{\alpha}) + \sigma_\tau \tau_i)\}^{-1}$, yielding the likelihood estimating equation as

$$\frac{\partial}{\partial \boldsymbol{\alpha}} \log \tilde{L}(\boldsymbol{\alpha}, \hat{\psi}(\cdot, \boldsymbol{\alpha})) = 0, \tag{5.26}$$

which may be solved by using the iterative equation

$$\hat{\boldsymbol{\alpha}}(r+1) = \hat{\boldsymbol{\alpha}}(r) - \left[\left(\frac{\partial^2}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} \log \tilde{L} \right)^{-1} \frac{\partial}{\partial \boldsymbol{\alpha}} \log \tilde{L} \right]_r, \tag{5.27}$$

where $[\cdot]_r$ indicates that the quantity in the square bracket is evaluated at $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}(r)$ obtained from the rth iteration. The solution of (5.26) is denoted as $\hat{\boldsymbol{\alpha}}_{SML}$. Note that the formulas for the first and second order derivatives for (5.26) and (5.27) are lengthy and cumbersome. For convenience, they are provided in Appendix A.4.

5.3 Asymptotic properties of the estimators of the SBDML model

5.3.1 Consistency of the SCQL estimator of $\psi(\cdot)$

The nonparametric function $\psi(z_{lu})$ in (5.1) has to be estimated for all $l = 1, \cdots, K$, and $u = 1, \cdots, n_l$. For convenience, in (5.12), we have shown the estimation of $\psi(z)$ for $z = z_{lu}$ for a selected value of l and u. Note that $\psi(z)$ cannot be estimated without knowing or estimating $\boldsymbol{\alpha}$. Thus, in (5.15), the estimate of $\psi(z)$ was denoted by $\hat{\psi}(z, \boldsymbol{\alpha})$. Now, a Taylor expansion of (5.12) about true $\psi(z)$ gives

$$\hat{\psi}(z, \boldsymbol{\alpha}) - \psi(z) \approx A_K + Q_K \,, \tag{5.28}$$

where $A_K = \frac{1}{B_K(z)} \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} h_{ij}(z)\, v_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\; [y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})] = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} a^{*}_{K,ij}(z)\; [y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})]$ with $B_K(z) = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} h_{ij}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\; v^{2}_{ijy_{i,j-1}}(z)$ and $h_{ij}(z)$ as the short abbreviation for $h_{ij}(\frac{z - z_{ij}}{b})$ defined in (5.14). Also in (5.28), $Q_K = \frac{1}{B_K(z)} \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} h_{ij}(z)\, v_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\; [p^{\dagger}_{ijy_{i,j-1}}(z_{ij}) - p^{\dagger}_{ijy_{i,j-1}}(z)] = \frac{1}{K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} a^{*}_{K,ij}(z)\; [p^{\dagger}_{ijy_{i,j-1}}(z_{ij}) - p^{\dagger}_{ijy_{i,j-1}}(z)]$. As n_i's are small and fixed, and K is large in the present longitudinal setup, we use $B = \lim_{K\to\infty} B_K$ for B_K involved in A_K. It then follows that A_K has zero mean and bounded variance, implying that $A_K = O_p(1/\sqrt{K})$ (Bishop, Fienberg, and Holland, 2007, Theorem 14.4-1). Furthermore, the following lemmas show that Q_K in (5.28) is in the order of $O(b^2)$.

Lemma 5.1. *Let* $\mathbf{x}_i = (\mathbf{x}_{i1}, \cdots, \mathbf{x}_{in_i})$, *then*

$$\mathrm{E}\left[\sum_{j=1}^{n_i} h_{ij}(z)\, v^{2}_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, (z_{ij} - z) \middle| \mathbf{x}_i \right] \;=\; O(b^2). \tag{5.29}$$

Proof. Let $\boldsymbol{z}_i = (z_{i1}, \cdots, z_{in_i})^{\top}$ and $q_{ij} = \Pr(y_{ij} = 1 | z_{i,j+1}, \boldsymbol{x}_i) = \Pr(y_{ij} = 1 | \boldsymbol{x}_i)$ because according to the model (5.1), the distribution of y_{ij} is independent of $z_{i,j+1}$. Then $q_{ij} = \int \mu_{ij} f_i(\boldsymbol{z}_i | \boldsymbol{x}_i) d\boldsymbol{z}_i$, where μ_{ij} is defined in (5.5), and $f_i(\boldsymbol{z}_i | \boldsymbol{x}_i)$ is the joint distribution of $\boldsymbol{z}_i$ conditional on $\boldsymbol{x}_i$. Also define

$$\begin{aligned} g_j(z_{ij}; \boldsymbol{\alpha}, z, \mathbf{x}_i) &= \mathrm{E}[v^{2}_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z) | z_{ij}, \mathbf{x}_i] \\ &= \sum_{y_{i,j-1}} v^{2}_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, q_{i,j-1}^{y_{i,j-1}} (1 - q_{i,j-1})^{1 - y_{i,j-1}} \end{aligned}$$

$$= g_j(\boldsymbol{\alpha}, z, \mathbf{x}_i)$$

because the conditional expectation is in fact independent of z_{ij}, and $h_j(z_{ij}; \mathbf{x}_i)$ be the pdf of z_{ij} conditional on $\mathbf{x}_i$, then

$$\begin{aligned}
& \mathrm{E}\left[\sum_{j=1}^{n_i} h_{ij}(z)\, v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, (z_{ij}-z) \middle| \mathbf{x}_i\right] \\
= & \sum_{j=1}^{n_i} \mathrm{E}_{z_{ij}} \left[h_{ij}(z)\,(z_{ij}-z)\, \mathrm{E}\left[v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, f_j(y_{i,j-1}|z_{ij}, \mathbf{x}_i) \middle| z_{ij}, \mathbf{x}_i\right] \middle| \mathbf{x}_i\right] \\
= & \sum_{j=1}^{n_i} \mathrm{E}_{z_{ij}} \left[h_{ij}(z)(z_{ij}-z) g_j(\boldsymbol{\alpha}, z, \mathbf{x}_i) | \mathbf{x}_i\right] \\
= & \sum_{j=1}^{n_i} \int h_{ij}(z) g_j(\boldsymbol{\alpha}, z, \mathbf{x}_i)(z_{ij}-z) h_j(z_{ij}; \mathbf{x}_i)\, dz_{ij}\,.
\end{aligned}$$

Then as $h_j(z_{ij}; \mathbf{x}_i) = h_j(z; \mathbf{x}_i) + O(z_{ij} - z)$, it follows that

$$\begin{aligned}
& \mathrm{E}\left[\sum_{j=1}^{n_i} h_{ij}(z)\, v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, (z_{ij}-z) \middle| \mathbf{x}_i\right] \\
= & \sum_{j=1}^{n_i} \int h_{ij}(z) \left[g_j(\boldsymbol{\alpha}, z, \mathbf{x}_i) h_j(z; \mathbf{x}_i)(z_{ij}-z) + O\left((z_{ij}-z)^2\right)\right] dz_{ij} \\
= & \sum_{j=1}^{n_i} g_j(\boldsymbol{\alpha}, z, \mathbf{x}_i) h_j(z; \mathbf{x}_i) \int h_{ij}(z)(z_{ij}-z)\, dz_{ij} + O(b^2) = O(b^2),
\end{aligned}$$

because $h_{ij}(z)$ is symmetric about z and $\int h_{ij}(z) O\left((z_{ij}-z)^2\right) dz_{ij}$ can be shown bounded in the order of $O(b^2)$. $\square$

Lemma 5.2.

$$Q_K \;=\; O(b^2). \tag{5.30}$$

Proof.

$$
\begin{aligned}
Q_K &\simeq \frac{1}{B(z)K}\sum_{i=1}^{K}\sum_{j=1}^{n_i} h_{ij}(z)\, v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, \psi'(z)\, [z_{ij}-z] \\
&= \frac{\psi'(z)}{B(z)K}\sum_{i=1}^{K}\Bigg\{\sum_{j=1}^{n_i} h_{ij}(z)\, v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, (z_{ij}-z) \\
&\qquad\qquad - \mathrm{E}\left[\sum_{j=1}^{n_i} h_{ij}(z)\, v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, (z_{ij}-z)\middle|\mathbf{x}_i\right]\Bigg\} \\
&\quad + \frac{\psi'(z)}{B(z)K}\sum_{i=1}^{K}\mathrm{E}\left[\sum_{j=1}^{n_i} h_{ij}(z)\, v^2_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, (z_{ij}-z)\middle|\mathbf{x}_i\right] = O(b^2),
\end{aligned}
$$

by result (5.29). For the first term, due to $h_{ij}(z)$, its variance is in the order of $O(b^2/K)$, so it is $O_p(b/\sqrt{K})$, which can be neglected. Thus we have shown that $Q_K = O(b^2)$. $\square$

Now we have

$$
A_K = O_p(1/\sqrt{K}) \quad \text{and} \quad Q_K = O(b^2)\,, \tag{5.31}
$$

yielding

$$
\hat{\psi}(z,\boldsymbol{\alpha}) - \psi(z) = A_K + O(b^2) = O_p(1/\sqrt{K}) + O(b^2)\,. \tag{5.32}
$$

That is, $\hat{\psi}(z,\boldsymbol{\alpha})$ is $\sqrt{K}$-consistent for $\psi(z)$ provided $Kb^4 \to 0$ as $K \to \infty$.

5.3.2 Asymptotic distribution of the SGQL estimator of $\boldsymbol{\alpha}$

Recall that $\boldsymbol{\alpha}$ represents the regression effects ($\boldsymbol{\beta}$), dynamic dependence parameter (θ) and the random effects variance (σ^2_τ). That is, $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma^2_\tau)^\top$, which was estimated in Section 5.2.2.1 by solving the SGQL estimating equation given by (5.18). Denote the estimating function (left side of (5.18)) involved in the SGQL equation as

$$\mathbf{D}_K(\boldsymbol{\alpha}) = \frac{1}{K}\sum_{i=1}^{K}\frac{\partial\tilde{\boldsymbol{\lambda}}_i^{\top}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}}\,\tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))\left[\mathbf{u}_i - \tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))\right], \qquad (5.33)$$

implying that $\mathbf{D}_K(\hat{\boldsymbol{\alpha}}_{SGQL}) = 0$. Now a linear Taylor expansion about true $\boldsymbol{\alpha}$ provides $\mathbf{D}_K(\boldsymbol{\alpha}) + \partial\mathbf{D}_K(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^{\top}\ (\hat{\boldsymbol{\alpha}}_{SGQL} - \boldsymbol{\alpha}) + o_p(1/\sqrt{K}) = 0$, yielding

$$\hat{\boldsymbol{\alpha}}_{SGQL} - \boldsymbol{\alpha} \;=\; \mathbf{F}_K^{-1}(\boldsymbol{\alpha})\,\mathbf{D}_K(\boldsymbol{\alpha}) + o_p(1/\sqrt{K})\,, \qquad (5.34)$$

where

$$\mathbf{F}_K(\boldsymbol{\alpha}) \;=\; -\frac{\partial\mathbf{D}_K(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^{\top}} = \frac{1}{K}\sum_{i=1}^{K}\frac{\partial\tilde{\boldsymbol{\lambda}}_i^{\top}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}}\,\tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))\frac{\partial\tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}^{\top}}.$$

Next, write $\mathbf{Z}_{1i} = \frac{\partial\tilde{\boldsymbol{\lambda}}_i^{\top}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}}\,\tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))$ and

$$\mathbf{Z}_{2ij} = \frac{1}{K}\sum_{i'=1}^{K}\sum_{j'=1}^{n_{i'}}\frac{\partial\tilde{\boldsymbol{\lambda}}_{i'}^{\top}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_{i'},\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}}\,\tilde{\boldsymbol{\Omega}}_{i'}^{-1}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_{i'},\boldsymbol{\alpha}))\,\frac{\partial\boldsymbol{\lambda}_{i'}(\boldsymbol{\alpha},\psi(\mathbf{z}_{i'}))}{\partial\psi(z_{i'j'})}\,a^*_{K,ij}(z_{i'j'})\,,$$

where $a^*_{K,ij}(z_{i'j'})$ is defined in (5.28). Because

$$\begin{aligned}\mathbf{D}_K(\boldsymbol{\alpha}) = \frac{1}{K}\sum_{i=1}^{K}\frac{\partial\tilde{\boldsymbol{\lambda}}_i^{\top}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))}{\partial\boldsymbol{\alpha}}\,\tilde{\boldsymbol{\Omega}}_i^{-1}(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))\,[\{\mathbf{u}_i - \boldsymbol{\lambda}_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))\} \\ + \left\{\boldsymbol{\lambda}_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i)) - \tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))\right\}\Big]\;,\end{aligned}$$

by using Taylor expansion of $\{\boldsymbol{\lambda}_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i)) - \tilde{\boldsymbol{\lambda}}_i(\boldsymbol{\alpha},\hat{\psi}(\mathbf{z}_i,\boldsymbol{\alpha}))\}$ with respect to $\psi(z_{i'j'})$ for all $i' = 1,\cdots,K$ and $j' = 1,\cdots,n_{i'}$, and applying (5.28) and (5.31), one may obtain

$$\mathbf{D}_K(\boldsymbol{\alpha}) = \frac{1}{K}\sum_{i=1}^{K}\left[\mathbf{Z}_{1i}\left(\mathbf{u}_i - \boldsymbol{\lambda}_i\right) - \sum_{j=1}^{n_i}\mathbf{Z}_{2ij}\left(y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right)\right]$$
$$+ O(b^2) + o_p(1/\sqrt{K}). \tag{5.35}$$

Denote $\mathbf{F} = \lim_{K\to\infty}\mathbf{F}_K$. It then follows from (5.34) by (5.35) that

$$\sqrt{K}\left\{\hat{\boldsymbol{\alpha}}_{SGQL} - \boldsymbol{\alpha}\right\} = \mathbf{F}^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left[\mathbf{Z}_{1i}\left(\mathbf{u}_i - \boldsymbol{\lambda}_i\right) - \sum_{j=1}^{n_i}\mathbf{Z}_{2ij}\left(y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right)\right]$$
$$+ O(\sqrt{Kb^4}) + o_p(1). \tag{5.36}$$

Next because $\mathrm{E}[\mathbf{U}_i - \boldsymbol{\lambda}_i] = 0$, $\mathrm{E}[Y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})] = 0$, and $\mathrm{Cov}[\mathbf{U}_i] = \boldsymbol{\Omega}_i$, by using Lindeberg-Feller central limit theorem (Amemiya, 1985, Theorem 3.3.6) for independent random variables with non-identical distributions, one obtains

$$\sqrt{K}\left\{\hat{\boldsymbol{\alpha}}_{SGQL} - \boldsymbol{\alpha} - O(b^2)\right\} \overset{D}{\to} N(0, \mathbf{V}_{SGQL}), \tag{5.37}$$

where

$$\mathbf{V}_{SGQL} = \mathbf{F}^{-1}\frac{1}{K}\left\{\sum_{i=1}^{K}\mathrm{Cov}\left[\mathbf{Z}_{1i}\left(\mathbf{U}_i - \boldsymbol{\lambda}_i\right) - \sum_{j=1}^{n_i}\mathbf{Z}_{2ij}\left(Y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right)\right]\right\}\mathbf{F}^{-1},$$

which can be estimated consistently by using

$$\hat{\mathbf{V}}_{SGQL} = \hat{\mathbf{F}}_K^{-1}(\hat{\boldsymbol{\alpha}}_{SGQL})\frac{1}{K}\left\{\sum_{i=1}^{K}\left[\mathbf{Z}_{1i}\left(\mathbf{U}_i - \boldsymbol{\lambda}_i\right) - \sum_{j=1}^{n_i}\mathbf{Z}_{2ij}\left(Y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right)\right]\right.$$
$$\left.\times\left[\mathbf{Z}_{1i}\left(\mathbf{U}_i - \boldsymbol{\lambda}_i\right) - \sum_{j=1}^{n_i}\mathbf{Z}_{2ij}\left(Y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right)\right]^{\top}\right\}\hat{\mathbf{F}}_K^{-1}(\hat{\boldsymbol{\alpha}}_{SGQL}).$$

Note that because $b \propto K^{-\nu}$, for $\sqrt{K}$-consistency of $\hat{\boldsymbol{\alpha}}_{SGQL}$, we need to have $Kb^4 \to 0$ as $K \to \infty$, which happens when $1/4 < \nu \leq 1/3$ (see Lin and Carroll, 2001, for example, for upper limit).

5.3.3 Asymptotic distribution of the SML estimator of $\boldsymbol{\alpha}$

Recall from Section 5.2.2.2 that the parameters of the model, i.e. $\boldsymbol{\alpha}$, were also estimated by using the SML approach, even though SGQL approach was found to be simpler. In this section, we now discuss the asymptotic properties of the SML estimator of $\boldsymbol{\alpha}$.

For true $\boldsymbol{\alpha}$, by using (5.1), we write $l(\boldsymbol{\alpha}, \psi(\cdot)) = \log L(\boldsymbol{\alpha}, \psi(\cdot)) = \sum_{i=1}^{K} l_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i)) = \sum_{j=1}^{n_i}[\mathbf{x}_{ij}^{\top}\boldsymbol{\beta} y_{ij} + \theta y_{i,j-1} y_{ij} + \psi(z_{ij}) y_{ij}] + \log J_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))$ because the individuals are independent. Here $J_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))$ is defined in (5.25). Recall from (5.26) that the likelihood estimate of $\boldsymbol{\alpha}$ was obtained by replacing $\psi(z_{ij})$ with its consistent estimate $\hat{\psi}(z_{ij}, \boldsymbol{\alpha})$. Applying a Taylor expansion over the likelihood equation gives

$$\sqrt{K}\left\{\hat{\boldsymbol{\alpha}}_{SML} - \boldsymbol{\alpha}\right\} \;=\; \mathbf{H}_K^{-1}\left\{\sqrt{K}\,\mathbf{C}_K\right\} + o_p(1), \tag{5.38}$$

where $\mathbf{C}_K = \frac{1}{K}\partial\tilde{l}(\boldsymbol{\alpha}, \hat{\psi}(\cdot\,, \boldsymbol{\alpha}))/\partial\boldsymbol{\alpha}$ and $\mathbf{H}_K = -\frac{1}{K}\partial^2\tilde{l}(\boldsymbol{\alpha}, \hat{\psi}(\cdot\,, \boldsymbol{\alpha}))/\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^{\top}$. Denote $\mathbf{H} = \lim_{K\to\infty}\mathbf{H}_K$. By (5.38), we write $\sqrt{K}\{\hat{\boldsymbol{\alpha}}_{SML} - \boldsymbol{\alpha}\} = \mathbf{H}^{-1}\{\sqrt{K}\,\mathbf{C}_K\} + o_p(1)$. By a further linear Taylor expansion of $\tilde{l}(\boldsymbol{\alpha}, \hat{\psi}(\cdot\,, \boldsymbol{\alpha}))$ about $\psi(z_{ij})$ for all $i = 1, \cdots, K$ and $j = 1, \cdots, n_i$, one may obtain $\tilde{l}(\boldsymbol{\alpha}, \hat{\psi}(\cdot\,, \boldsymbol{\alpha})) \simeq l(\boldsymbol{\alpha}, \psi(\cdot)) + \sum\limits_{i=1}^{K}\sum\limits_{j=1}^{n_i} \frac{\partial l(\boldsymbol{\alpha}, \psi(\cdot))}{\partial\psi(z_{ij})}\,[\hat{\psi}(z_{ij}, \boldsymbol{\alpha}) - \psi(z_{ij})]$, yielding

$$\mathbf{C}_K \approx \frac{1}{K}\sum_{i=1}^{K}\frac{\partial l_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))}{\partial\boldsymbol{\alpha}} + \mathbf{C}_{1K} + \frac{1}{K}\sum_{i=1}^{K}\sum_{j=1}^{n_i}\frac{\partial\hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\frac{\partial l_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))}{\partial\psi(z_{ij})}\,, \tag{5.39}$$

where $\mathbf{C}_{1K} = \frac{1}{K}\sum\limits_{i=1}^{K}\sum\limits_{j=1}^{n_i}\mathbf{Z}_{3ij}(y_{i,j-1})[y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})] + O(b^2)$ with $\mathbf{Z}_{3ij}(y_{i,j-1}) = \frac{1}{K}\sum\limits_{i'=1}^{K}\sum\limits_{j'=1}^{n_{i'}}$

$\frac{\partial^2 l_{i'}(\boldsymbol{\alpha},\psi(\mathbf{z}_{i'}))}{\partial\boldsymbol{\alpha}\partial\psi(z_{i'j'})}$ $a^*_{K,ij}(z_{i'j'})$, where $a^*_{K,ij}(z_{i'j'})$ is defined in (5.28). Next by using the law of large numbers, we can neglect the dependence of $\mathbf{Z}_{3ij}$ on $y_{i'j'}$'s, and retain only its dependence on $y_{i,j-1}$ contained in $a^*_{K,ij}(z_{i'j'})$. Also by the law of large numbers, in (5.39), we can neglect the dependence of $\frac{\partial\hat{\psi}(z_{ij},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}$ on $\mathbf{x}_{ij}$'s and y_{ij}'s, and regard it as a function of z_{ij} and $\boldsymbol{\alpha}$ only. Then (5.38) becomes

$$\begin{aligned}\sqrt{K}\left\{\hat{\boldsymbol{\alpha}}_{SML}-\boldsymbol{\alpha}\right\} \approx\ & \mathbf{H}^{-1}\frac{1}{\sqrt{K}}\sum_{i=1}^{K}\left\{\frac{\partial l_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))}{\partial\boldsymbol{\alpha}}+\sum_{j=1}^{n_i}\mathbf{Z}_{3ij}(y_{i,j-1})\left[y_{ij}-p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right]\right. \\ & \left.+\sum_{j=1}^{n_i}\frac{\partial\hat{\psi}(z_{ij},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\frac{\partial l_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))}{\partial\psi(z_{ij})}\right\}+O(\sqrt{Kb^4}). \qquad (5.40)\end{aligned}$$

Because $\mathrm{E}[\partial l_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))/\partial\boldsymbol{\alpha}]=0$ and $\mathrm{E}[\partial l_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))/\partial\psi(z_{ij})]=0$ for all i and j, by Lindeberg-Feller central limit theorem (Amemiya, 1985, Theorem 3.3.6) we can obtain

$$\sqrt{K}\left\{\hat{\boldsymbol{\alpha}}_{SML}-\boldsymbol{\alpha}-O(b^2)\right\}\xrightarrow{D}N(0,\mathbf{V}_{SML}), \qquad (5.41)$$

where

$$\begin{aligned}\mathbf{V}_{SML}=\mathbf{H}^{-1}\frac{1}{K}\sum_{i=1}^{K}\mathrm{Cov}&\left\{\frac{\partial l_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))}{\partial\boldsymbol{\alpha}}+\sum_{j=1}^{n_i}\mathbf{Z}_{3ij}(y_{i,j-1})\left[y_{ij}-p^{\dagger}_{ijy_{i,j-1}}(z_{ij})\right]\right. \\ &\left.+\sum_{j=1}^{n_i}\frac{\partial\hat{\psi}(z_{ij},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\frac{\partial l_i(\boldsymbol{\alpha},\psi(\mathbf{z}_i))}{\partial\psi(z_{ij})}\right\}\mathbf{H}^{-1},\end{aligned}$$

which can be estimated consistently by using

$$\hat{\mathbf{V}}_{SML}=\hat{\mathbf{H}}_K^{-1}\left\{\frac{1}{K}\sum_{i=1}^{K}\hat{\mathbf{G}}_i\hat{\mathbf{G}}_i^{\top}\right\}\hat{\mathbf{H}}_K^{-1},$$

where

$$\hat{\mathbf{G}}_i = \Big\{\frac{\partial l_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))}{\partial \boldsymbol{\alpha}} + \sum_{j=1}^{n_i} \mathbf{Z}_{3ij}(y_{i,j-1})[y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z_{ij})]$$
$$+ \sum_{j=1}^{n_i} \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \frac{\partial l_i(\boldsymbol{\alpha}, \psi(\mathbf{z}_i))}{\partial \psi(z_{ij})}\Big\}\Big|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}, \psi(\mathbf{z}_i)=\hat{\psi}(\mathbf{z}_i, \hat{\boldsymbol{\alpha}})}.$$

It follows from (5.41) that $\hat{\boldsymbol{\alpha}}_{SML}$ is biased for $\boldsymbol{\alpha}$ unless $\sqrt{K}\, O(b^2) \to 0$. Hence for $\sqrt{K}$-consistency of $\hat{\boldsymbol{\alpha}}_{SML}$, we need to have $Kb^4 \to 0$ as $K \to \infty$, which gives $K \frac{1}{K^{4\nu}} = \frac{1}{K^{4\nu-1}} \to 0$, and hence $1/4 < \nu \leq 1/3$ (see Lin and Carroll, 2001, for example, for upper limit).

We remark that in obtaining SGQL and SML estimates of $\boldsymbol{\alpha}$, we have used $\hat{\psi}(z, \boldsymbol{\alpha})$ as a consistent estimate for $\psi(z)$. However, when $\boldsymbol{\alpha}$ is estimated, this estimate in turn becomes $\hat{\psi}(z, \hat{\boldsymbol{\alpha}})$. Note that no matter whether $\boldsymbol{\alpha}$ is estimated by SGQL or SML approach, as long as ν satisfies $1/4 < \nu \leq 1/3$,

$$\sqrt{K}\, O\left((\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^2\right) \;=\; \frac{1}{\sqrt{K}}\, O\left(\left(\sqrt{K}\,(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})\right)^2\right) \;=\; \frac{1}{\sqrt{K}}\, O_p(1) \;=\; o_p(1).$$

It then follows that $\hat{\psi}(z, \hat{\boldsymbol{\alpha}})$ is also $\sqrt{K}$-consistent for $\psi(z)$ under the condition $Kb^4 \to 0$, or equivalently, $\frac{1}{4} < \nu \leq \frac{1}{3}$. That is,

$$\begin{aligned}
\sqrt{K}\left\{\hat{\psi}(z, \hat{\boldsymbol{\alpha}}) - \psi(z)\right\} &= \sqrt{K}\left\{\hat{\psi}(z, \hat{\boldsymbol{\alpha}}) - \hat{\psi}(z, \boldsymbol{\alpha})\right\} + \sqrt{K}\left\{\hat{\psi}(z, \boldsymbol{\alpha}) - \psi(z)\right\} \\
&= \frac{1}{\sqrt{K}} \sum_{i=1}^{K} \sum_{j=1}^{n_i} a^*_{K,ij}(z)\left[y_{ij} - p^{\dagger}_{ijy_{i,j-1}}(z)\right] \\
&\quad + \frac{\partial \hat{\psi}(z, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\top}} \sqrt{K}\, \{\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\} + O(\sqrt{K b^4}) + o_p(1)\,.
\end{aligned}$$

5.4 A simulation study

In Chapter 4, specifically in Section 4.2.4, we conducted a simulation study examining the performance of the SML estimation approach under the semi-parametric binary fixed model. However, because the SML approach becomes complicated for the semi-parametric mixed model in this chapter, we have discussed the SGQL approach mainly for the estimation of the main parameters $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma_\tau^2)^\top$. This approach is simpler than the SML approach for such mixed models. The large sample properties of the SGQL estimator of $\boldsymbol{\alpha}$ was discussed in Section 5.3.2. In this section, we now conduct a simulation study to examine the small/finite sample performance of the SGQL estimator. For the sake of completeness, we also include the SML approach in this study.

As far as the design is concerned, we use 2 fixed covariates as in Chapters 3 and 4 (Sections 3.4, 4.1.4 and 4.2.4). However, we write these covariates again as follows to accommodate a general cluster size n_i. More specifically, we will use $n_i = 4$, 6 and 10. These notations will reflect the covariates of the previous chapters when $n_i = 4$.

$$x_{ij1} = \begin{cases} \frac{1}{2} & \text{for } i = 1, \ldots, 25 \text{ and } j \leq n_i/2 \\ 1 & \text{for } i = 1, \ldots, 25 \text{ and } n_i/2 < j \leq n_i \\ -\frac{1}{2} & \text{for } i = 26, \ldots, 75 \text{ and } j \leq \left\lfloor \frac{n_i}{3} \right\rfloor \\ 0 & \text{for } i = 26, \ldots, 75 \text{ and } \left\lfloor \frac{n_i}{3} \right\rfloor < j \leq n_i - \left\lfloor \frac{n_i}{3} \right\rfloor \\ \frac{1}{2} & \text{for } i = 26, \ldots, 75 \text{ and } j > n_i - \left\lfloor \frac{n_i}{3} \right\rfloor \\ \frac{j}{2n_i} & \text{for } i = 76, \ldots, 100 \text{ and } j = 1, \cdots, n_i, \end{cases} \tag{5.42}$$

and

$$x_{ij2} = \begin{cases} \frac{j-(n_i+1)/2}{2n_i} & \text{for } i = 1, \ldots, 50 \text{ and } j = 1, \cdots, n_i \\ 0 & \text{for } i = 51, \ldots, 100 \text{ and } j \leq n_i/2 \\ \frac{1}{2} & \text{for } i = 51, \ldots, 100 \text{ and } j > n_i/2, \end{cases} \tag{5.43}$$

where $\lfloor q \rfloor$ denotes the largest integer $\leq q$.

Next, for the selection of the nonparametric function in secondary covariates $z_{ij} \sim U[j - 0.5, j + 0.5]$, we choose the same function as in the past chapters. That is,

$$\psi(z_{ij}) \quad = \quad 0.3 + 0.2\left(z_{ij} - \frac{n_i + 1}{2}\right) + 0.05\left(z_{ij} - \frac{n_i + 1}{2}\right)^2.$$

With regard to the selection of the parameters of the model, we choose the same fixed regression and dynamic independence parameters as in Section 4.2.4. That is, $\beta_1 = \beta_2 = 0.5$ and $\theta = 1.0, -1.0, -3.0$. However, for the additional random effect variance parameter σ_τ^2, we choose

$$\sigma_\tau^2 \quad = \quad 0.5, 1.0, 2.0 \text{ and } 3.0. \tag{5.44}$$

Our main objective in this section is to examine the effect of σ_τ^2 in the estimation of the other parameters and nonparametric function of the mixed model.

For data generation, as in the mixed model for count data discussed in Section 3.4.2, we first generate random effects $\tau_i^* \sim N(0, \sigma_\tau^2)$ for a given value of σ_τ^2 from (5.44). We then use (5.1) to generate data for all selected values of $\boldsymbol{\beta}$ and θ.

The estimation approaches given in Section 5.2 require computing statistic means over τ_i as in Eqs. (5.5) and (5.7). To calculate such statistic means over τ_i by averaging over a large sample of τ_i is too time-consuming to be practical. Instead in this work we applied a binomial approximation approach as described in Appendix A.2.

5.4.1 Estimation performance for $\psi(\cdot)$ and $\boldsymbol{\alpha} = (\boldsymbol{\beta}^\top, \theta, \sigma_\tau^2)^\top$ for various σ_τ^2

We now examine the performance of the proposed SGQL and SML approaches discussed in Section 5.2 for the estimation of $\psi(z_{ij})$, and the parameters $\boldsymbol{\beta}$, θ and σ_τ^2 for $\sigma_\tau^2 = 0.5, 1, 2, 3$, respectively. We consider 4 ($n_i = 4$, $i = 1, \cdots, K$) repeated binary responses from each of $K = 100$ independent individuals. All estimates (simulated mean, SM) along with their simulated standard errors (SSE) and mean square errors (MSE) are obtained based on 1000 simulations. The results for parameters $\boldsymbol{\beta}$, θ and σ_τ^2 are provided in Tables 5.1 and 5.2. The SCQL estimates for the function $\psi(\cdot)$ are displayed in Figs. 5.1 - 5.4. Here the bandwidth in $\psi(\cdot)$ estimation is chosen as $b = c_0 K^{-1/5}$ with $c_0 = \sigma_z \approx \left(\frac{z \text{ range}}{4}\right)$ as mentioned in Section 4.2.2.1. For the integrations over τ_i, we used the binomial approximation proposed in Appendix A.2, with number of trials equal to 5. Whenever the estimated $\hat{\sigma}_\tau^2$ was negative, it was set to 10^{-8}, and the iterative algorithm continued until convergence.

The results from Tables 5.1 and 5.2 show that the SGQL and SML approaches give almost the same SMs, SSEs and MSEs for the estimation of all the parameters $\boldsymbol{\beta}$, θ and σ_τ^2 in all the parameter combinations considered, and the estimates of the nonparametric function in Figs. 5.1 - 5.4 with parameters estimated by these 2 approaches also coincide with each other, indicating the strength and advantage of the proposed SGQL approach, as SGQL is considerably easier to implement, and spends much less computer time in getting convergent estimates. Because of the closeness of the estimation results from these two approaches, we will concentrate on only SML when discussing the performance of these two approaches.

As seen in Tables 5.1 and 5.2, the estimates are in general good agreement with

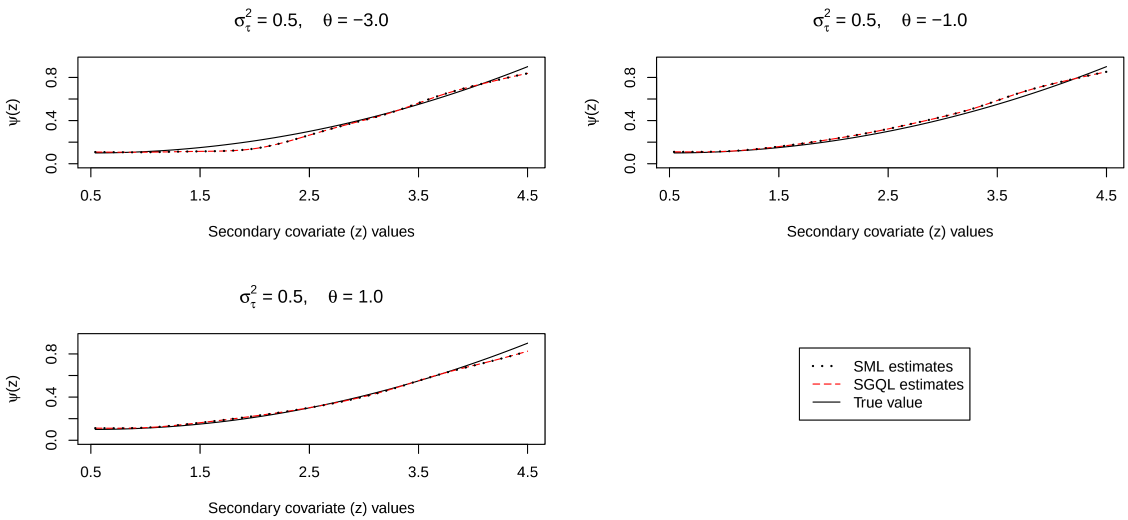

Figure 5.1: Estimated nonparametric function for SBDML model. Using $\sigma_\tau^2 = 0.5$ and $n_i = 4$ for all i.

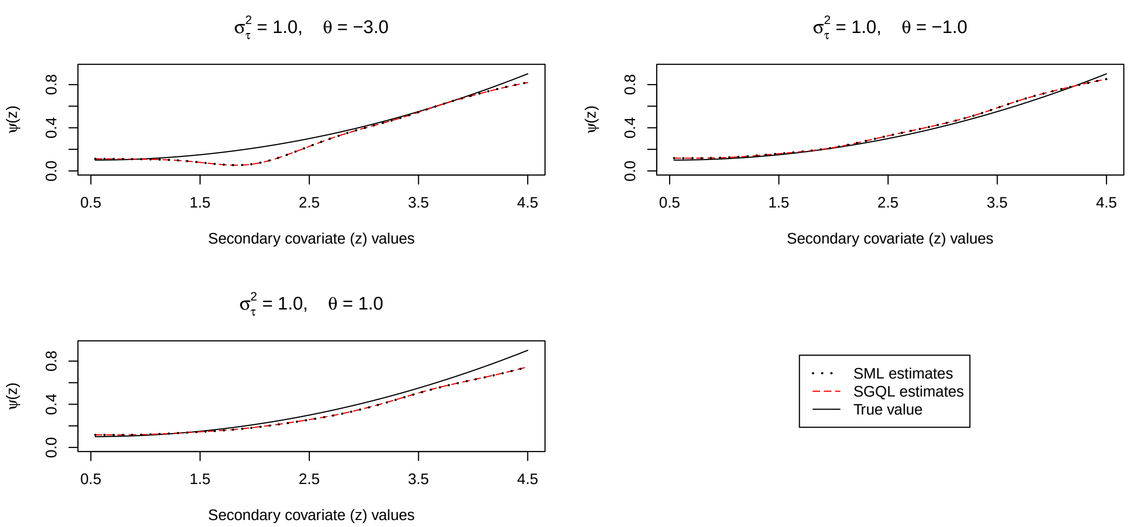

Figure 5.2: Estimated nonparametric function for SBDML model. Using $\sigma_\tau^2 = 1.0$ and $n_i = 4$ for all i.

the corresponding true values of the parameters. To be specific, the SML approach appears to produce almost unbiased estimates for the dynamic dependence parameter. For example, $\theta = 1.0$ is estimated as 1.0144 and $\theta = -3.0$ is estimated as -3.0516

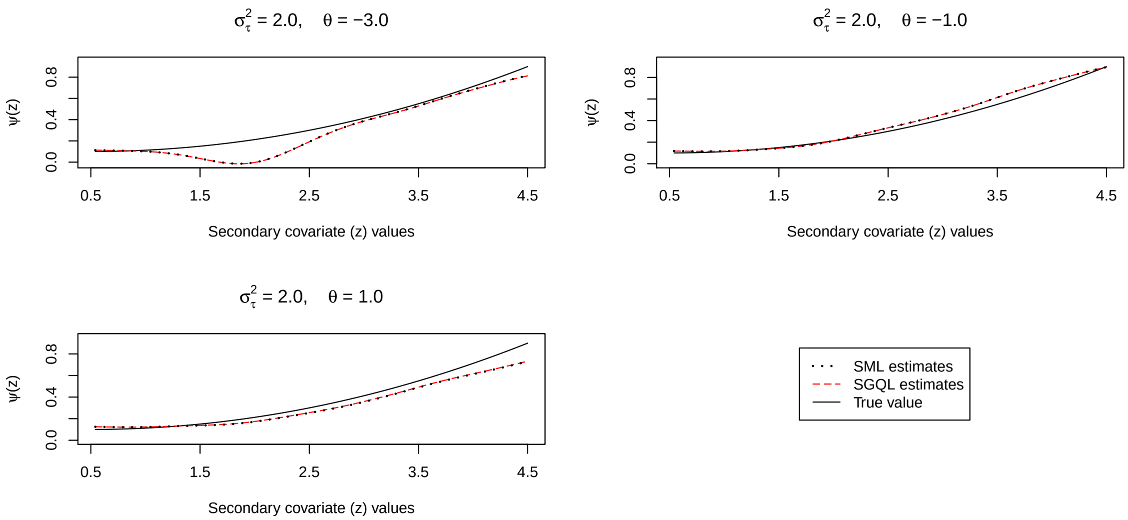

Figure 5.3: Estimated nonparametric function for SBDML model. Using $\sigma_\tau^2 = 2.0$ and $n_i = 4$ for all i.

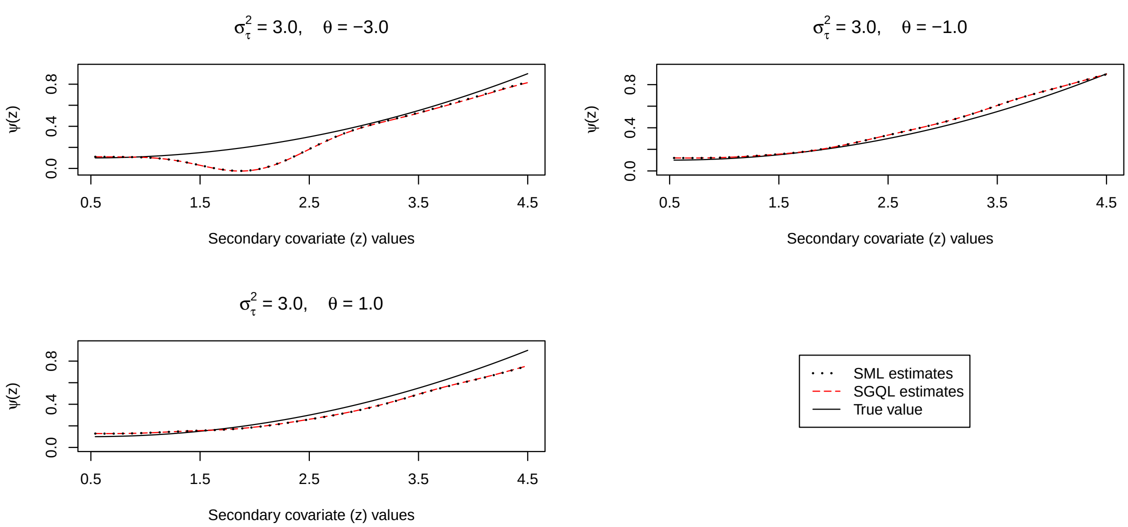

Figure 5.4: Estimated nonparametric function for SBDML model. Using $\sigma_\tau^2 = 3.0$ and $n_i = 4$ for all i.

when $\sigma_\tau^2 = 3.0$. As far as the estimation of regression effects is concerned, when σ_τ^2 is as small as 0.5 and 1.0, similar to the situation for SBDL model in Section 4.2.4, $\boldsymbol{\beta}$ estimates are less biased when dynamic dependence is negative. For example, for

$\sigma_\tau^2 = 0.5$, the $\beta_2 = 0.5$ estimate is 0.5354 when $\theta = 1.0$, but the estimate is found to be 0.5288 when $\theta = -1.0$, and 0.5042 when $\theta = -3.0$. However, when σ_τ^2 is as large as 2.0 and 3.0, situation becomes more complicated. For example, when $\sigma_\tau^2 = 2.0$, the $\beta_2 = 0.5$ estimate is 0.5370 for $\theta = 1.0$, gets as better as 0.4940 for $\theta = -1.0$, and then becomes 0.4556 for $\theta = -3.0$, which is worse than both results for $\theta = 1.0$ and -1.0. Further note that the estimates for β_2 are in general considerably worse than those for β_1, which can be because there are more variations in covariate x_{ij1} than in x_{ij2}. With regard to the estimation of random effect variance, in all the cases considered, σ_τ^2 estimate is less biased with smaller standard error and mean square error when dynamic dependence goes to negative. For example, when $\sigma_\tau^2 = 3.0$, for $\theta = 1.0, -1.0$ and -3.0, the σ_τ^2 estimates are 3.3317, 3.0270 and 3.0119 with SSEs 2.0502, 1.2153 and 1.1147, and MSEs 4.3092, 1.4761 and 1.2415, respectively. Notice that when $\theta = 1.0$ and $\sigma_\tau^2 = 0.5$, the estimate of σ_τ^2 is 0.6012 with a large bias. This is because the negative estimates of σ_τ^2 are set to 10^{-8}, causing the distribution of $\hat{\sigma}_\tau^2$ right tilted. In practice, sample size is usually quite large, and the dynamic dependence parameter θ is also smaller than 1.0, then the proposed estimation approaches will be adequate in fitting the model.

Next, Figs. 5.1 - 5.4 show that the SCQL approach estimates the true nonparametric curve well when σ_τ^2 and $|\theta|$ are not too large. The estimated curves almost coincide with the true curve when $\theta = -1.0$ for all σ_τ^2 values from 0.5 to 3.0. When $\theta = 1.0$, the curve estimate is good for $\sigma_\tau^2 = 0.5$, but shows slight underestimation of $\psi(\cdot)$ value in the right part of the range of secondary covariate z for large σ_τ^2 values such as 1.0, 2.0 and 3.0. When absolute value of θ is large, such as -3.0, but σ_τ^2 is as small as 0.5, the estimated nonparametric function still closely follows the true curve. However, when σ_τ^2 is also large, that is, when $\sigma_\tau^2 = 1.0, 2.0$ and 3.0, there appears considerable bias in the range approximately from 1.1 to 2.9, and this bias increases as

σ_τ^2 gets larger. The bias in nonparametric function estimation comes from the second term in (5.28), or more specifically, from

$$p^\dagger_{ijy_{i,j-1}}(z_{ij}) - p^\dagger_{ijy_{i,j-1}}(z)\,. \tag{5.45}$$

So for the covariate and nonparametric function configuration in this simulation study, when $\theta = -3.0$ and σ_τ^2 are large, (5.45) is negatively biased from 0 in the second time point with $z_{i2} \in [1.5, 2.5]$ and $\mathbf{x}_{i2} = (x_{i21}, x_{i22})^\top$ given by (5.42) and (5.43). This bias will decrease to 0 when sample size gets larger, according to the asymptotic results given in Section 5.3. Also, in practice it should be rare for a large random effect variance and a large-magnitude negative dynamic dependence parameter to happen together, even though they may happen separately, so a satisfying estimation for the nonparametric function will be anticipated for the application of our models and estimation approaches to real data analysis.

Table 5.1: Simulated means (SMs), simulated standard errors (SSEs) and mean square errors (MSEs) of SGQL and SML estimates of the parameters $\boldsymbol{\beta}$, θ and σ_τ^2, under SBDML model (5.1) for selected parameter values with $K = 100$, $n_i = 4$, and 1000 simulations.

Methods	Quantity	β_1	β_2	θ	σ_τ^2
	True Value	0.5	0.5	1.0	0.5
SGQL	SM	0.5334	0.5350	1.0062	0.6032
	SSE	0.4439	1.0386	0.4721	0.6825
	MSE	0.1980	1.0789	0.2227	0.4760
SML	SM	0.5332	0.5354	1.0062	0.6012
	SSE	0.4433	1.0373	0.4716	0.6753
	MSE	0.1974	1.0761	0.2222	0.4658
	True Value	0.5	0.5	-1.0	0.5
SGQL	SM	0.5133	0.5290	-1.0364	0.5245
	SSE	0.3827	0.8490	0.3669	0.3747
	MSE	0.1465	0.7209	0.1358	0.1409
SML	SM	0.5131	0.5288	-1.0365	0.5246
	SSE	0.3829	0.8488	0.3664	0.3741
	MSE	0.1466	0.7206	0.1354	0.1404
	True Value	0.5	0.5	-3.0	0.5
SGQL	SM	0.5073	0.5048	-3.0800	0.5110
	SSE	0.4057	0.9416	0.4538	0.3700
	MSE	0.1645	0.8858	0.2122	0.1369
SML	SM	0.5071	0.5042	-3.0811	0.5121
	SSE	0.4060	0.9424	0.4539	0.3698
	MSE	0.1647	0.8872	0.2124	0.1367
	True Value	0.5	0.5	1.0	1.0
SGQL	SM	0.5188	0.5431	1.0509	1.0902
	SSE	0.4805	1.1288	0.5048	0.9719
	MSE	0.2310	1.2747	0.2572	0.9517
SML	SM	0.5180	0.5426	1.0517	1.0848
	SSE	0.4798	1.1249	0.5037	0.9552
	MSE	0.2303	1.2659	0.2562	0.9187
	True Value	0.5	0.5	-1.0	1.0
SGQL	SM	0.5111	0.5476	-1.0394	1.0425
	SSE	0.4370	0.9442	0.3883	0.5241
	MSE	0.1909	0.8929	0.1521	0.2763
SML	SM	0.5107	0.5474	-1.0393	1.0421
	SSE	0.4374	0.9447	0.3880	0.5245
	MSE	0.1913	0.8939	0.1519	0.2766
	True Value	0.5	0.5	-3.0	1.0
SGQL	SM	0.5085	0.4930	-3.0539	1.0092
	SSE	0.4595	1.0124	0.4817	0.5148
	MSE	0.2110	1.0239	0.2347	0.2648
SML	SM	0.5084	0.4919	-3.0562	1.0117
	SSE	0.4599	1.0110	0.4803	0.5131
	MSE	0.2113	1.0212	0.2336	0.2632

Table 5.2: Table 5.1 continued.

Methods	Quantity	β_1	β_2	θ	σ_τ^2
	True Value	0.5	0.5	1.0	2.0
SGQL	SM	0.5130	0.5366	1.0449	2.1779
	SSE	0.5661	1.2877	0.5470	1.4623
	MSE	0.3204	1.6578	0.3009	2.1677
SML	SM	0.5116	0.5370	1.0473	2.1628
	SSE	0.5645	1.2866	0.5453	1.4246
	MSE	0.3185	1.6549	0.2993	2.0540
	True Value	0.5	0.5	−1.0	2.0
SGQL	SM	0.4821	0.4949	−1.0166	2.0361
	SSE	0.5064	1.0859	0.4157	0.8409
	MSE	0.2565	1.1779	0.1729	0.7077
SML	SM	0.4813	0.4940	−1.0160	2.0343
	SSE	0.5062	1.0870	0.4150	0.8387
	MSE	0.2563	1.1804	0.1723	0.7039
	True Value	0.5	0.5	−3.0	2.0
SGQL	SM	0.4875	0.4570	−3.0407	1.9938
	SSE	0.5205	1.1654	0.4819	0.7670
	MSE	0.2708	1.3586	0.2337	0.5878
SML	SM	0.4875	0.4556	−3.0447	1.9988
	SSE	0.5197	1.1643	0.4839	0.7683
	MSE	0.2700	1.3561	0.2359	0.5896
	True Value	0.5	0.5	1.0	3.0
SGQL	SM	0.5184	0.5965	1.0147	3.3345
	SSE	0.6270	1.4056	0.5810	2.0506
	MSE	0.3931	1.9831	0.3374	4.3126
SML	SM	0.5196	0.5981	1.0144	3.3317
	SSE	0.6266	1.4091	0.5809	2.0502
	MSE	0.3926	1.9932	0.3373	4.3092
	True Value	0.5	0.5	−1.0	3.0
SGQL	SM	0.4930	0.5291	−1.0236	3.0353
	SSE	0.5674	1.1750	0.4350	1.2315
	MSE	0.3216	1.3802	0.1896	1.5163
SML	SM	0.4909	0.5285	−1.0215	3.0270
	SSE	0.5661	1.1754	0.4334	1.2153
	MSE	0.3203	1.3809	0.1881	1.4761
	True Value	0.5	0.5	−3.0	3.0
SGQL	SM	0.5040	0.4527	−3.0459	3.0025
	SSE	0.5891	1.2594	0.5248	1.0998
	MSE	0.3467	1.5868	0.2773	1.2084
SML	SM	0.5033	0.4477	−3.0516	3.0119
	SSE	0.5887	1.2583	0.5283	1.1147
	MSE	0.3463	1.5845	0.2815	1.2415

5.4.2 Naive estimation of $\boldsymbol{\beta}$, θ and $\psi(\cdot)$ (ignoring σ_τ^2)

Note that when data are generated under the present SBDML model (5.1) following the aforementioned specifications, but one ignores the presence of random effect in the model and makes an attempt to estimate the parameters $\boldsymbol{\beta}$ and θ as well as the nonparametric function $\psi(\cdot)$ by treating the data as though they were generated from the SBDL model (4.39)−(4.40), the estimates are bound to be biased. We examine the performance of such naive SML (NSML) estimators by repeating the data generation 1000 times and computing the simulated mean (SM), simulated standard error (SSE), and mean square error (MSE) of the NSML estimates for $\boldsymbol{\beta}$ and θ, and the simulated mean of the SCQL estimates of the function $\psi(\cdot)$. The parameter values and their simulated estimates are shown in Table 5.3. The true nonparametric function, and its estimates by assuming SBDL and SBDML models are displayed in Fig. 5.5.

Table 5.3: Simulated means (SMs), simulated standard errors (SSEs) and mean square errors (MSEs) of NSML estimates of parameters $\boldsymbol{\beta}$ and θ, with data generated under SBDML model for selected parameter values with $K = 100$, $n_i = 4$, and 1000 simulations.

Methods	Quantity	β_1	β_2	θ	σ_τ^2
	True Value	0.5	0.5	1.0	1.0
NSML	SM	0.3541	0.3366	1.6777	
	SSE	0.3706	0.9015	0.3256	
	MSE	0.1585	0.8387	0.5652	
	True Value	0.5	0.5	−1.0	1.0
NSML	SM	0.3779	0.4226	−0.2711	
	SSE	0.3480	0.7705	0.2685	
	MSE	0.1359	0.5991	0.6032	
	True Value	0.5	0.5	−3.0	1.0
NSML	SM	0.3966	0.4500	−2.1750	
	SSE	0.3799	0.8471	0.3227	
	MSE	0.1549	0.7194	0.7846	

As expected, the results in Table 5.3 show that the estimates of $\boldsymbol{\beta}$ and θ are highly

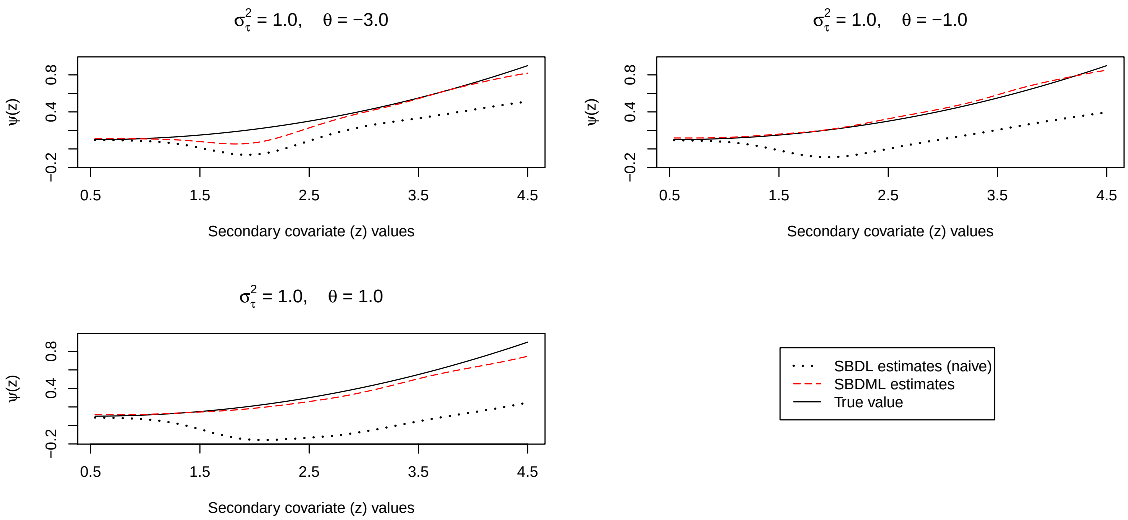

Figure 5.5: Estimated nonparametric function for SBDML model.

biased. For example, when $\sigma_\tau^2 = 1.0$, for the true regression parameter $\boldsymbol{\beta} = (0.5, 0.5)^\top$ and dynamic dependence parameter $\theta = 1.0$, the estimated values of $\boldsymbol{\beta}$ and θ are found to be $(0.3541, 0.3366)^\top$ and 1.6777, respectively. The naive estimates of the function $\psi(\cdot)$ in Fig. 5.5 also show large bias on almost the whole range of $\psi(\cdot)$. Clearly all of these naive estimates computed by ignoring the random effect are useless, and hence one must take the random effect into account in estimating these regression and dynamic dependence parameters, and the nonparametric function. This will require the consistent estimation of the random effect variance σ_τ^2, which was discussed in Section 5.2.

Note that the estimating equations are constructed by minimizing the generalized distance between the estimated individual means and the data. As a result, sometimes wrong models can still give close estimates of the true means. To further investigate the difference in estimating the true means between fitting SBDML and SBDL models, we plotted in Figs. 5.6, 5.7 and 5.8 the true means (black solid lines), the means computed with estimated parameter values and nonparametric function by fitting

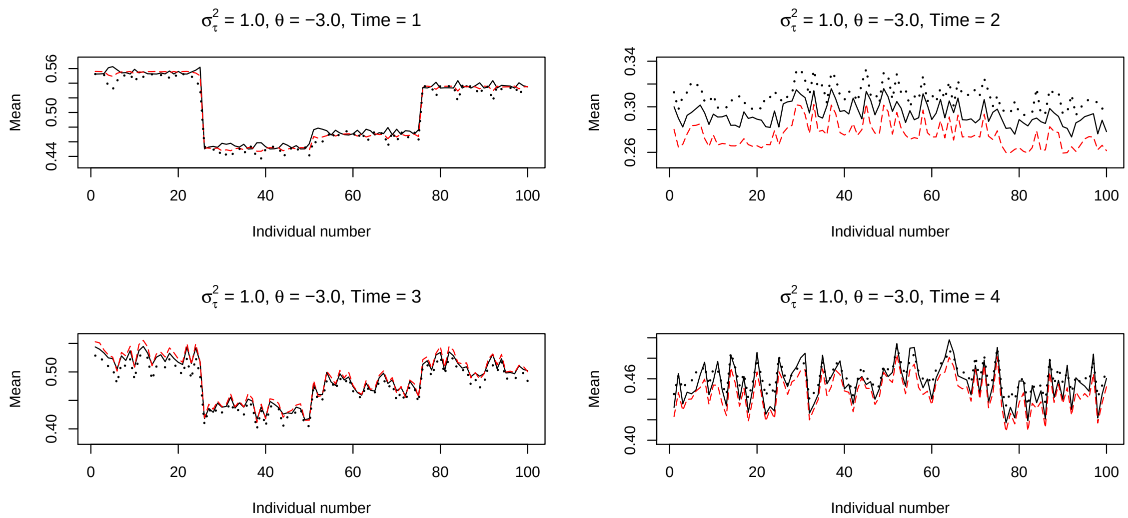

Figure 5.6: (Color online) True and estimated means for SBDML model. The black solid lines are the true means, the red dashed lines are the means estimated by fitting SBDML model, and the black dotted lines are the means estimated by fitting SBDL model.

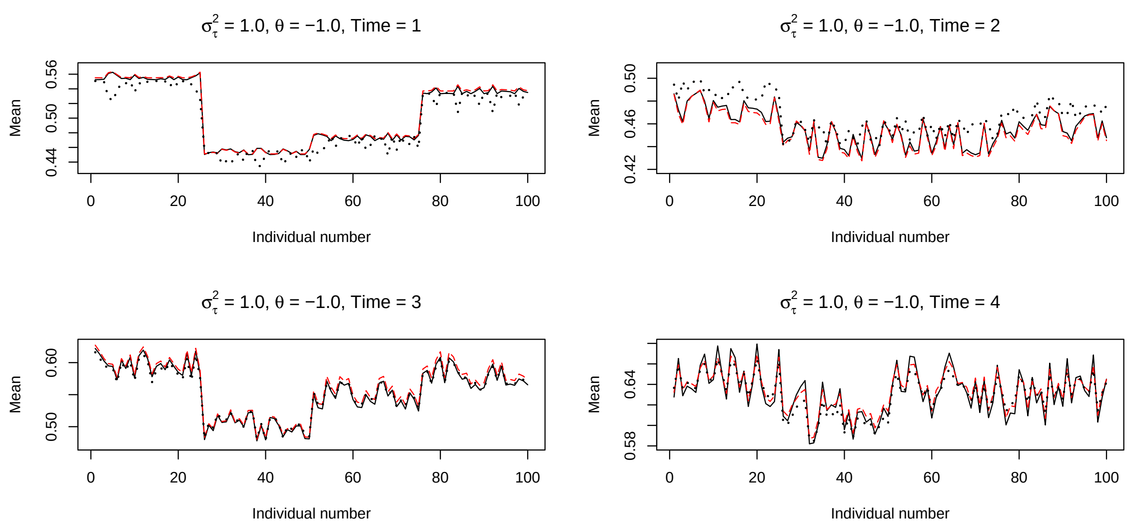

Figure 5.7: (Color online) True and estimated means for SBDML model. The black solid lines are the true means, the red dashed lines are the means estimated by fitting SBDML model, and the black dotted lines are the means estimated by fitting SBDL model.

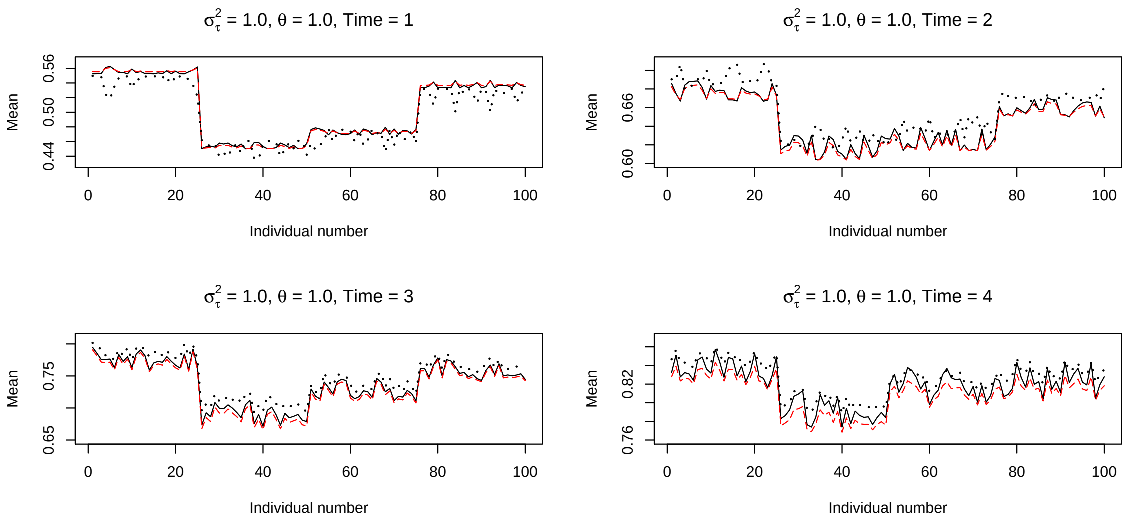

Figure 5.8: (Color online) True and estimated means for SBDML model. The black solid lines are the true means, the red dashed lines are the means estimated by fitting SBDML model, and the black dotted lines are the means estimated by fitting SBDL model.

SBDML model with SML approach (red dashed lines), as given in Table 5.1 and Fig. 5.2, and the means computed with parameters and nonparametric function estimated by fitting SBDL model with SML approach (black dotted lines), as given in Table 5.3 and Fig. 5.5. The figures show that the means estimated under true SBDML model are in general considerably closer to the true means, as compared to the means estimated under SBDL model, especially for the first several time points, and the mean estimation under SBDML model remains good from starting time 1 to ending time 4, while the mean estimation under SBDL model improves as time increases. For example, in Fig. 5.8, the means estimated under SBDML model almost overlap with the true mean for time 1, 2 and 3, while the means estimated under SBDL model frequently oscillate away from the true means, showing comparatively not small differences between the true means and the estimated means for most individuals. Only at the last time point, time 4, the mean estimates from the two models have similar distances from the true mean. Also clear in this figure is that the estimates from fitting SBDL model become better with time increase.

5.4.3 Estimation performance of the proposed approaches for a single data set

Notice that the simulation results presented in Figs. 5.1 − 5.4 and Tables 5.1 and 5.2 are average performance of the estimates based on 1000 simulated data sets. However, in practice, usually only 1 data set is available. So it is necessary for us to explore the performance of the proposed model and estimation approaches on a single data set. Particularly, since the semi-parametric fixed models and SGEE estimating approach for binary panel data have been studied before this work (see Lin and Carroll, 2001, for example), while the corresponding mixed models are left untouched, we would like to explore here the performance of SBDML model on a single data set, and compare it with that of SBDL model. To be specific, we generated a single data set from SBDML model (5.1) with $\beta_1 = \beta_2 = 0.5$, $\theta = 0.2$, and $\sigma_\tau^2 = 1.0$ and 3.0, estimated the parameters and nonparametric function by applying SBDL and SBDML models with SML approach, and compared the estimation results. Note that the SML approach for SBDL model is just the SML approach for SBDML model with the random effect variance $\sigma_\tau^2 = 0$. Here the series length for the ith individual, n_i, is chosen to be 6 and 10. The regression covariates are still given by (5.42) and (5.43), and the secondary covariate (z_{ij}) generation and the nonparametric function definition still follow the design described at the start of this simulation section.

Table 5.4 gives the estimated parameter values. For all the four cases, β_1 estimates with SBDML model are all considerably better than those with SBDL model. This is also almost true for θ estimation, except the case with $\sigma_\tau^2 = 1.0$ and $n_i = 6$, where the θ estimates from the two models have nearly the same distance from the true value

(0.6193 for SBDL, and 0.6669 for SBDML). Only for β_2, SBDL gives better estimates than SBDML in the four cases. This becomes understandable when referring to Tables 5.1 and 5.2, where, for a similar covariate configuration, and the same sample size K, β_2 estimates always have much larger standard errors than β_1 and θ estimates, that is, β_2 estimate bears more randomness, and hence there can be more cases that wrong models give better estimates of β_2 than the true model. Whereas, in Table 5.4, the overall performance of SBDML is better than that of SBDL.

Table 5.4: The parameter estimates for SBDML model. Here one data set is generated with SBDML model for each parameter value combination. The parameters are estimated using SBDL and SBDML models.

Methods	Quantity	β_1	β_2	θ	σ_τ^2	n_i
	True Value	0.5	0.5	0.2	1.0	6
SBDL	Estimation	0.0354	1.1961	0.8193		
SBDML	Estimation	0.3489	1.7027	-0.4669	2.3876	
	True Value	0.5	0.5	0.2	3.0	6
SBDL	Estimation	-0.1581	0.4606	1.8889		
SBDML	Estimation	0.0983	0.2602	0.1798	4.3419	
	True Value	0.5	0.5	0.2	1.0	10
SBDL	Estimation	0.4980	0.7246	0.9590		
SBDML	Estimation	0.5005	0.8130	0.6285	0.4149	
	True Value	0.5	0.5	0.2	3.0	10
SBDL	Estimation	0.2179	-0.2521	2.0486		
SBDML	Estimation	0.5535	-0.6137	0.6018	3.1533	

The estimated nonparametric functions, along with the true ones, are displayed in Figs. 5.9 and 5.10. By checking these graphs, one may obtain the following observations. The SBDML estimates of the function $\psi(\cdot)$ are fairly closer to the true ones than the SBDL estimates. When random effect variance σ_τ^2 gets larger, SBDL estimates of $\psi(\cdot)$ become worse. This is reasonable since when σ_τ^2 gets larger, the difference between the wrong SBDL and true SBDML models becomes larger. On the contrary, SBDML estimates of $\psi(\cdot)$ may not become worse (as in Fig. 5.9), or even becomes better (as in Fig. 5.10), as σ_τ^2 gets larger. To explain this phenomenon, we can look at SBDML model (5.1). When estimating the function $\psi(\cdot)$, the random effect $\sigma_\tau \tau_i$ behaves as a confounder, and the difference between the smooth function $\psi(\cdot)$ and the randomly varying $\sigma_\tau \tau_i$ grows as σ_τ^2 increases, which makes it easier for the SCQL algorithm to abstract the smooth $\psi(\cdot)$ from the random $\sigma_\tau \tau_i$.

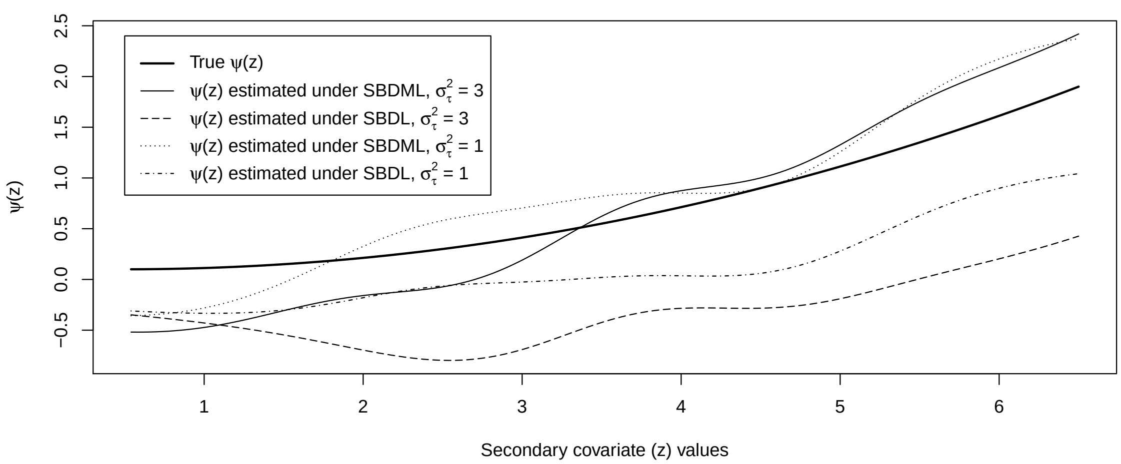

Figure 5.9: The true and estimated non-parametric functions for a single data set generated from the SBDML model (5.1) with $n_i = 6, \beta_1 = \beta_2 = 0.5, \theta = 0.2$, and $\sigma_\tau^2 = 1.0$ and 3.0. The estimates are obtained by fitting SBDL and SBDML models.

To futher compare the performances of SBDL and SBDML for single data sets, we also plotted the average estimated means from SBDL and SBDML models, as well as the average response values, at each time point j for the cases of $(n_i = 6, \sigma_\tau^2 = 3.0)$ and $(n_i = 10, \sigma_\tau^2 = 3.0)$ in Figs. 5.11 and 5.12, respectively. The average response values and the average estimated means are calculated by

$$\begin{aligned}\bar{y}_j &= \frac{\sum_{i=1}^K y_{ij}}{K} \\ \bar{\hat{\mu}}_j &= \frac{\sum_{i=1}^K \hat{\mu}_{ij}(\hat{\boldsymbol{\alpha}}, \hat{\psi}(\mathbf{z}_i, \hat{\boldsymbol{\alpha}}))}{K},\end{aligned} \tag{5.46}$$

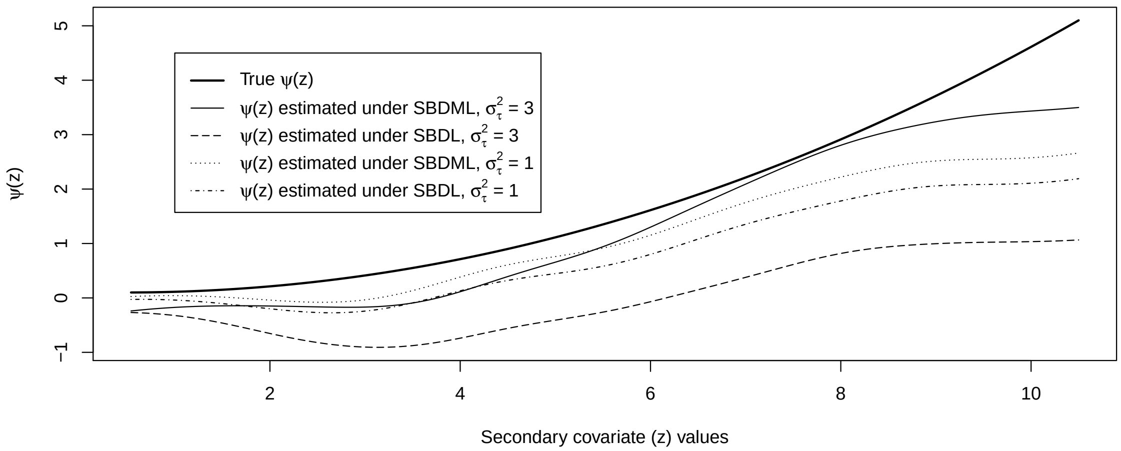

Figure 5.10: The true and estimated non-parametric functions for a single data set generated from the SBDML model (5.1) with $n_i = 10, \beta_1 = \beta_2 = 0.5, \theta = 0.2$, and $\sigma_\tau^2 = 1.0$ and 3.0. The estimates are obtained by fitting SBDL and SBDML models.

respectively. Notice that the average estimated means based on SBDML model are considerably closer to the average response values for the first several time points, especially for the first 2 time points, as compared to those based on SBDL model. While for larger time points, the average estimated means from the two models are close to each other, and to the average response values. The difference between the estimates from the 2 models at larger time points increases as σ_τ^2 becomes larger. In practice, the individual series length n_i can be small, which is especially true for unbalanced data sets, where n_i can be quite small for parts of individuals. Under such circumstances, the better performance of SBDML model for initial time points will show its advantage over SBDL model.

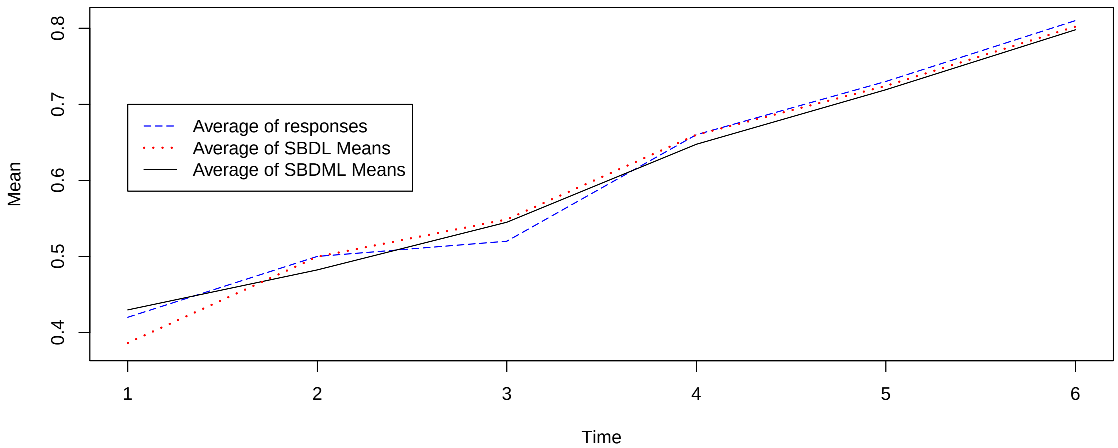

Figure 5.11: The average of the estimated means and the average of response values at each longitudinal time point for a single data set generated from the SBDML model (5.1) with $n_i = 6, \beta_1 = \beta_2 = 0.5, \theta = 0.2$ and $\sigma_\tau^2 = 3.0$. The estimates are obtained by fitting SBDL and SBDML models.

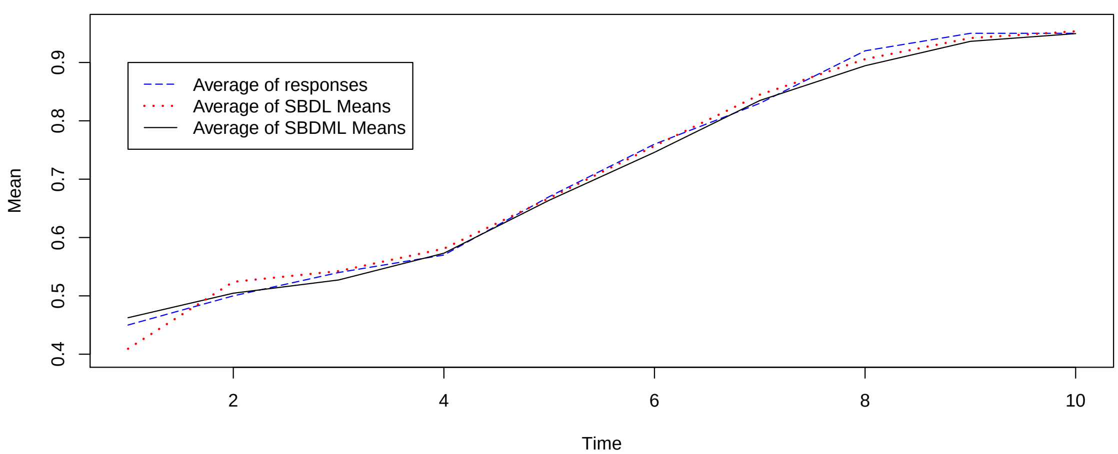

Figure 5.12: The average of the estimated means and the average of response values at each longitudinal time point for a single data set generated from the SBDML model (5.1) with $n_i = 10, \beta_1 = \beta_2 = 0.5, \theta = 0.2$ and $\sigma_\tau^2 = 3.0$. The estimates are obtained by fitting SBDL and SBDML models.

Chapter 6

Conclusion

Longitudinal data analysis for discrete such as count and binary data has been an important research topic over the last three decades. Because of the efficiency drawbacks (Crowder, 1995; Sutradhar and Das, 1999; Sutradhar, 2011, Chapter 6) of the GEE (generalized estimating equation) approach (Liang and Zeger, 1986) in dealing with such data, there have been alternative studies using a parametric class of autocorrelations, where the GQL (generalized quasi-likelihood) (Sutradhar, 2003, 2010, 2011) approach is used for inferences about the main parameters of the model. The GEE approach later on was extended to deal with longitudinal semi-parametric models (Severini and Staniswalis, 1994, Lin and Carroll, 2001). Recently, this type of semi-parametric model for linear data was studied by Warriyar and Sutradhar (2014) using parametric correlations based GQL approach. Furthermore, Sutradhar et al. (2016) have extended the longitudinal semi-parametric models for linear data to the count data setup. We have summarized these studies in Chapter 2.

Note that the aforementioned studies on semi-parametric models for longitudinal data were confined to fixed effect cases. However, because of the importance of mixed regression effects in longitudinal setup (Manski, 1987, Wooldridge, 1999, Honoré and

Kyriazidou, 2000, Sutradhar and Bari, 2007, Sutradhar et al., 2008, 2010), we have extended these longitudinal mixed models to the semi-parametric setup for count data in Chapter 3. Some of the results dealing with longitudinal semi-parametric mixed models for count data are also available in Zheng and Sutradhar (2016).

In Chapter 4, we studied the longitudinal semi-parametric fixed effect models for binary data. We have considered two specific correlation models, namely SLDCP (semi-parametric linear dynamic conditional probability) and SBDL (semi-parametric binary dynamic logit) models. In studying these models, we have noticed an undesirable feature of the SGEE approaches used by Lin and Carroll (2001) (see Section 4.1.4.1). Specifically, we found that UNS (unstructured) based SGEE approach used by these and other authors produces less efficient estimates than simpler approaches based on the independence assumption. The proposed SGQL inference technique under the SLDCP model does not suffer from this type of inefficiency problems. Furthermore, the BDL model (Sutradhar and Farrell, 2007; Sutradhar, 2011, Chapter 7) was extended to the semi-parametric setup in Section 4.2. Detailed asymptotic properties and finite sample results are studied for both SLDCP and SBDL models. These models and SGQL and SML (semi-parametric maximum likelihood) inference techniques were illustrated by using the binary infectious disease data (see also Lin and Carroll, 2001).

Finally, in Chapter 5, we have studied the analysis of longitudinal binary data by using semi-parametric longitudinal mixed models. More specifically, we have modeled such binary data using the SBDML (semi-parametric binary dynamic mixed logit) models. Step by step estimation for the nonparametric function and parameters of the models were given. Both finite sample and asymptotic properties of the estimators were also discussed in details.

We remark that even though the estimates for the nonparametric function and variance components of the random effects were found reasonable for moderately large variance component in the mixed model setup, these estimates specifically for large variance component could be improved using some bias corrections to the present inference techniques. However, this type of bias correction is deferred to future studies. Furthermore, in longitudinal studies, there can be missing responses causing difficulties for the inferences about nonparametric function and regression and other parameters. This is also beyond the scope of the present thesis. Finally, our future studies can also include, but not limited to, comparing our method of bandwidth selection, namely $b = \sigma_z K^{-1/5}$, with cross-validation and generalized cross-validation, simulation studies regarding the model misspecification with respect to the choice of the kernel bandwidth and the distribution of random effects, and application of a parametric piecewise constant model for the effects of the secondary covariate instead of the nonparametric model.

Appendix A

Computational details for SCQL, SGQL and SML estimating equations under SBDML model in Chapter 5

A.1 Higher order (conditional and unconditional) moments computation

Recall from Section 5.3.2 that the SGQL approach requires the computational formulas for higher order such as third and fourth order moments. For convenience, we provide the formulas for these moments as follows:

Lemma A.1. *The third and fourth order conditional moments of the ith individual,* *given random effect τ_i, are given by*

$$\begin{aligned}
\mathrm{E}[Y_{ij}Y_{ik}Y_{im}|\tau_i] &= \mu^*_{im}\mu^*_{ij}\left(1-\mu^*_{ij}\right)\left[\prod_{l=j+1}^{k}\left(p^*_{il1}-p^*_{il0}\right)\right] + \left(1-\mu^*_{ik}\right)\mu^*_{ij}\left(1-\mu^*_{ij}\right)\left[\prod_{l=j+1}^{m}\left(p^*_{il1}-p^*_{il0}\right)\right] \\
&\quad + \mu^*_{ij}\mu^*_{ik}\left(1-\mu^*_{ik}\right)\left[\prod_{l=k+1}^{m}\left(p^*_{il1}-p^*_{il0}\right)\right] + \mu^*_{ij}\mu^*_{ik}\mu^*_{im} = \delta^*_{ijkm}\,, \qquad say,
\end{aligned} \tag{A.1}$$

with $j < k < m$, and

$$\begin{aligned}
\mathrm{E}[Y_{ij}Y_{ik}Y_{im}Y_{in}|\tau_i] &= \mu^*_{ij}\mu^*_{im}\left(1-\mu^*_{ij}\right)\left(1-\mu^*_{im}\right)\left[\prod_{l=j+1}^{k}\left(p^*_{il1}-p^*_{il0}\right)\right]\left[\prod_{l=m+1}^{n}\left(p^*_{il1}-p^*_{il0}\right)\right] \\
&+ \mu^*_{ij}\mu^*_{ik}\mu^*_{im}\left(1-\mu^*_{im}\right)\left[\prod_{l=m+1}^{n}\left(p^*_{il1}-p^*_{il0}\right)\right] + \mu^*_{ij}\mu^*_{ik}\left(1-\mu^*_{ik}\right)\left(1-\mu^*_{im}\right)\left[\prod_{l=k+1}^{n}\left(p^*_{il1}-p^*_{il0}\right)\right] \\
&+ \mu^*_{ij}\mu^*_{ik}\mu^*_{in}\left(1-\mu^*_{ik}\right)\left[\prod_{l=k+1}^{m}\left(p^*_{il1}-p^*_{il0}\right)\right] + \mu^*_{ij}\left(1-\mu^*_{ij}\right)\left(1-\mu^*_{ik}\right)\left(1-\mu^*_{im}\right)\left[\prod_{l=j+1}^{n}\left(p^*_{il1}-p^*_{il0}\right)\right] \\
&+ \mu^*_{ij}\mu^*_{in}\left(1-\mu^*_{ij}\right)\left(1-\mu^*_{ik}\right)\left[\prod_{l=j+1}^{m}\left(p^*_{il1}-p^*_{il0}\right)\right] + \mu^*_{ij}\mu^*_{im}\mu^*_{in}\left(1-\mu^*_{ij}\right)\left[\prod_{l=j+1}^{k}\left(p^*_{il1}-p^*_{il0}\right)\right] \\
&+ \mu^*_{ij}\mu^*_{ik}\mu^*_{im}\mu^*_{in} = \phi^*_{ijkmn}\,, \qquad say,
\end{aligned} \tag{A.2}$$

with $j < k < m < n$, respectively.

Proof. By applying formula (5.9) recursively, the third and fourth order conditional central moments can be obtained as: for $j < k < m$,

$$\mathrm{E}[\left(Y_{ij}-\mu^*_{ij}\right)\left(Y_{ik}-\mu^*_{ik}\right)\left(Y_{im}-\mu^*_{im}\right)|\tau_i] = \left(1-2\mu^*_{ik}\right)\mu^*_{ij}\left(1-\mu^*_{ij}\right)\left[\prod_{l=j+1}^{m}\left(p^*_{il1}-p^*_{il0}\right)\right], \tag{A.3}$$

and for $j < k < m < n$,

$$\begin{aligned}
&\mathrm{E}[\left(Y_{ij} - \mu^*_{ij}\right)\left(Y_{ik} - \mu^*_{ik}\right)\left(Y_{im} - \mu^*_{im}\right)\left(Y_{in} - \mu^*_{in}\right)|\tau_i] \\
&= \left(1 - 2\mu^*_{im}\right)\left(1 - 2\mu^*_{ik}\right)\mu^*_{ij}\left(1 - \mu^*_{ij}\right)\left[\prod_{l=j+1}^{n}\left(p^*_{il1} - p^*_{il0}\right)\right] \\
&+ \mu^*_{ij}\left(1 - \mu^*_{ij}\right)\mu^*_{im}\left(1 - \mu^*_{im}\right)\left[\prod_{l=m+1}^{n}\left(p^*_{il1} - p^*_{il0}\right)\right]\left[\prod_{l=j+1}^{k}\left(p^*_{il1} - p^*_{il0}\right)\right]. \qquad \text{(A.4)}
\end{aligned}$$

Then applying the second order moment based result (5.6) (see also Farrell and Sutradhar, 2006, for example), Equations (A.1) and (A.2) follow from expanding left hand sides of (A.3) and (A.4). □

Lemma A.2. *The third and fourth order unconditional moments are given by*

$$\begin{aligned}
\delta_{ijkm} \equiv \mathrm{E}[Y_{ij}Y_{ik}Y_{im}] &= \int_{-\infty}^{\infty} \delta^*_{ijkm}(\tau_i)\phi(\tau_i)d\tau_i \simeq \frac{1}{N}\sum_{w=1}^{N}\delta^*_{ijkm}(\tau_{iw}), \quad \textit{and} \\
\phi_{ijkmn} \equiv \mathrm{E}[Y_{ij}Y_{ik}Y_{im}Y_{in}] &= \int_{-\infty}^{\infty} \phi^*_{ijkmn}(\tau_i)\phi(\tau_i)d\tau_i \simeq \frac{1}{N}\sum_{w=1}^{N}\phi^*_{ijkmn}(\tau_{iw}), \qquad \text{(A.5)}
\end{aligned}$$

respectively.

A.2 Unconditional moments using binomial approximation

For the SGQL estimation of parameters in Section 5.2.2.1, we need to calculate the unconditional moments such as δ_{ijkm} and ϕ_{ijkmn} as in (A.5), which are cumbersome to simplify the normal integral over τ_i's. This type of normal integral is also needed in the SML estimation of parameters given in Section 5.2.2.2 and the SCQL estimation of the nonparametric function given in Section 5.2.1. A simpler way to compute the desired normal integral is to approximate it by a binomial approximation (see Ten Have and Morabia (1999), eqn. (7), for example). For example, one may compute μ_{ij} as

$$\mu_{ij} \simeq \sum_{\nu_i=0}^{V} \mu^*_{ij}(\tau_i) \begin{pmatrix} V \\ \nu_i \end{pmatrix} \left(\frac{1}{2}\right)^{\nu_i} \left(\frac{1}{2}\right)^{V-\nu_i}, \tag{A.6}$$

where for a known reasonably big V such as $V = 5$, $\nu_i \sim$ binomial$(V, 1/2)$, and hence it has a relation to τ_i as $\tau_i = \frac{\nu_i - V(1/2)}{\sqrt{V(1/2)(1/2)}}$. In the same way, we can calculate other moments as

$$\begin{aligned} \lambda_{ijk} &\simeq \sum_{\nu_i=0}^{V} \lambda^*_{ijk}(\tau_i) \begin{pmatrix} V \\ \nu_i \end{pmatrix} \left(\frac{1}{2}\right)^{\nu_i} \left(\frac{1}{2}\right)^{V-\nu_i}, \\ \delta_{ijkm} &\simeq \sum_{\nu_i=0}^{V} \delta^*_{ijkm}(\tau_i) \begin{pmatrix} V \\ \nu_i \end{pmatrix} \left(\frac{1}{2}\right)^{\nu_i} \left(\frac{1}{2}\right)^{V-\nu_i}, \\ \phi_{ijkmn} &\simeq \sum_{\nu_i=0}^{V} \phi^*_{ijkmn}(\tau_i) \begin{pmatrix} V \\ \nu_i \end{pmatrix} \left(\frac{1}{2}\right)^{\nu_i} \left(\frac{1}{2}\right)^{V-\nu_i}. \end{aligned} \tag{A.7}$$

A.3 Derivatives of the estimated nonparametric function $\hat{\psi}(z, \boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$

Lemma A.3. *The first order derivatives of the estimated nonparametric $\hat{\psi}(z, \boldsymbol{\alpha})$ with respect to $\boldsymbol{\beta}$, θ and σ^2_τ are given by*

$$\frac{\partial \hat{\psi}(z, \boldsymbol{\beta}, \theta, \sigma^2_\tau)}{\partial \boldsymbol{\beta}} = -\frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z)\, \sigma^{-2}_{ijy_{i,j-1}}(z)\, U_{ij}\, \mathbf{x}_{ij}}{\sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z)\, \sigma^{-2}_{ijy_{i,j-1}}(z)\, U_{ij}}, \tag{A.8}$$

$$\frac{\partial \hat{\psi}(z, \boldsymbol{\beta}, \theta, \sigma_\tau^2)}{\partial \theta} = -\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\,\sigma_{ijy_{i,j-1}}^{-2}(z)\,U_{ij}\,y_{i,j-1}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\,\sigma_{ijy_{i,j-1}}^{-2}(z)\,U_{ij}}\,, \quad \textit{and} \tag{A.9}$$

$$\frac{\partial \hat{\psi}(z, \boldsymbol{\beta}, \theta, \sigma_\tau^2)}{\partial \sigma_\tau^2} = -\frac{1}{2\sigma_\tau}\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\,\sigma_{ijy_{i,j-1}}^{-2}(z)\,R_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\,\sigma_{ijy_{i,j-1}}^{-2}(z)\,U_{ij}} \tag{A.10}$$

respectively, where

$$\begin{aligned}
U_{ij} &= \Big\{q_{ijy_{i,j-1}}(z)\,\sigma_{ijy_{i,j-1}}(z) - \big(1 - 2p_{ijy_{i,j-1}}(z)\big)\,v_{ijy_{i,j-1}}^2(z)\Big\}\,\big(y_{ij} - p_{ijy_{i,j-1}}(z)\big) \\
&\quad - v_{ijy_{i,j-1}}^2(z)\,\sigma_{ijy_{i,j-1}}(z)\,, \quad \textit{and} \\
R_{ij} &= \big\{e_{ijy_{i,j-1}}(z)\,\sigma_{ijy_{i,j-1}}(z) - v_{ijy_{i,j-1}}(z)\,\big[1 - 2p_{ijy_{i,j-1}}(z)\big]\,a_{ijy_{i,j-1}}(z)\big\}\,\big(y_{ij} - p_{ijy_{i,j-1}}(z)\big) \\
&\quad - v_{ijy_{i,j-1}}(z)\,a_{ijy_{i,j-1}}(z)\,\sigma_{ijy_{i,j-1}}(z)
\end{aligned}$$

with

$$\begin{aligned}
a_{ijy_{i,j-1}}(z) &= \int_{-\infty}^{\infty} p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\,\Big[1 - p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\Big]\,\tau_i\phi(\tau_i)d\tau_i\,, \\
e_{ijy_{i,j-1}}(z) &= \int_{-\infty}^{\infty} \Big(1 - 2p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\Big)\,p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\,\Big(1 - p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\Big)\,\tau_i\phi(\tau_i)d\tau_i\,, \\
q_{ijy_{i,j-1}}(z) &= \int_{-\infty}^{\infty} \Big(1 - 2p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\Big)\,p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\,\Big(1 - p_{ijy_{i,j-1}}^*(\boldsymbol{\alpha}, \psi(z), \tau_i)\Big)\,\phi(\tau_i)d\tau_i\,.
\end{aligned}$$

Proof. Derivative of the estimation equation (5.12) with respect to $\boldsymbol{\beta}$ gives

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\,\Big\{q_{ijy_{i,j-1}}(z)\sigma_{ijy_{i,j-1}}(z) - \big(1 - 2p_{ijy_{i,j-1}}(z)\big)\,v_{ijy_{i,j-1}}^2(z)\Big\}\,\sigma_{ijy_{i,j-1}}^{-2}(z)\,\big(y_{ij} - p_{ijy_{i,j-1}}(z)\big)$$

$$\times \left(\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right) - \sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z)\, v^{2}_{ijy_{i,j-1}}(z) \left(\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right) = 0\,. \tag{A.11}$$

We then solve for $\partial\hat{\psi}(z,\boldsymbol{\alpha})/\partial\boldsymbol{\beta}$ to obtain (A.8).

Similarly, by taking the derivative of the estimation equation (5.12) with respect to θ and σ^2_τ, and solving for $\partial\hat{\psi}(z,\boldsymbol{\beta},\theta,\sigma^2_\tau)/\partial\theta$ and $\partial\hat{\psi}(z,\boldsymbol{\beta},\theta,\sigma^2_\tau)/\partial\sigma^2_\tau$ respectively, we obtain (A.9) and (A.10). □

Lemma A.4. *The second order derivatives of the estimated nonparametric $\hat{\psi}(z,\boldsymbol{\alpha})$ with respect to $\boldsymbol{\beta}$, θ and σ^2_τ are given by*

$$\frac{\partial^2 \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \simeq -\frac{1}{m_K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z) b_{ij} \left(\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right) \left(\mathbf{x}^\top_{ij} + \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \boldsymbol{\beta}^\top} \right),$$

$$\frac{\partial^2 \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \theta^2} \simeq -\frac{1}{m_K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z) b_{ij} \left(y_{i,j-1} + \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \theta} \right)^2, \ \textit{and}$$

$$\frac{\partial^2 \hat{\psi}(z,\boldsymbol{\alpha})}{\partial {\sigma^2_\tau}^2} \simeq$$

$$\begin{aligned} \frac{1}{m_K} \sum_{i=1}^{K} \sum_{j=1}^{n_i} w_{ij}(z) \Bigg\{ & 2\sigma^{-2}_{ijy_{i,j-1}}(z) v_{ijy_{i,j-1}}(z) \left(1 - 2p_{ijy_{i,j-1}}(z)\right) \left(\frac{a_{ijy_{i,j-1}}(z)}{2\sigma_\tau} + v_{ijy_{i,j-1}}(z) \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \sigma^2_\tau} \right)^2 \\ & - \sigma^{-1}_{ijy_{i,j-1}}(z) \left[2 \left(\frac{e_{ijy_{i,j-1}}(z)}{2\sigma_\tau} + q_{ijy_{i,j-1}}(z) \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \sigma^2_\tau} \right) \left(\frac{a_{ijy_{i,j-1}}(z)}{2\sigma_\tau} + v_{ijy_{i,j-1}}(z) \frac{\partial \hat{\psi}(z,\boldsymbol{\alpha})}{\partial \sigma^2_\tau} \right) \right. \\ & \left. + v_{ijy_{i,j-1}}(z) \left(g_{ijy_{i,j-1}}(z) - \frac{a_{ijy_{i,j-1}}(z)}{4\sigma^3_\tau} \right) \right] \Bigg\} \end{aligned} \tag{A.12}$$

respectively, where

$$b_{ij} = 3v_{ijy_{i,j-1}}(z)\, q_{ijy_{i,j-1}}(z)\, \sigma^{-1}_{ijy_{i,j-1}}(z) - 2\sigma^{-2}_{ijy_{i,j-1}}(z) \left[1 - 2p_{ijy_{i,j-1}}(z)\right] v^{3}_{ijy_{i,j-1}}(z)\,,$$

$$g_{ijy_{i,j-1}}(z) = \int_{-\infty}^{\infty} \left(1 - 2p^{*}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z),\tau_i)\right) p^{*}_{ijy_{i,j-1}}(\boldsymbol{\alpha},\psi(z),\tau_i)$$

$$\left(1 - p^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z), \tau_i)\right)\left(\frac{\tau_i}{2\sigma_\tau} + \frac{\partial\hat{\psi}(z,\boldsymbol{\alpha})}{\partial\sigma^2_\tau}\right)^2 \phi(\tau_i)d\tau_i\,, \;\; \textit{and}$$

$$m_K = \sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)\,\sigma^{-1}_{ijy_{i,j-1}}(z)\,v^2_{ijy_{i,j-1}}(z). \tag{A.13}$$

Proof. By taking the derivative of Eq. (A.11) with respect to $\boldsymbol{\beta}^\top$, and neglecting the quantities containing $\{y_{ij} - p_{ijy_{i,j-1}}(z)\}$, we obtain

$$-\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z)b_{ij}\left(\mathbf{x}_{ij} + \frac{\partial\hat{\psi}(z,\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}}\right)\left(\mathbf{x}^\top_{ij} + \frac{\partial\hat{\psi}(z,\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}^\top}\right) - m_K\frac{\partial^2\hat{\psi}(z,\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top} \simeq 0\,,$$

leading to the result for $\partial^2\hat{\psi}(z,\boldsymbol{\alpha})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^\top$ as in (A.12). The other two second order derivatives in (A.12) can also be obtained using the same procedure. □

A.4 Derivatives of log-likelihood under SBDML model

Lemma A.5. *The components of $\partial \log \tilde{L}/\partial\boldsymbol{\alpha}$ for (5.26) under the SBDML model have the forms:*

$$\frac{\partial \log \tilde{L}}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[y_{ij} - \frac{A_{ij}}{J_i}\right]\left[\mathbf{x}_{ij} + \frac{\partial\hat{\psi}(z_{ij},\boldsymbol{\alpha})}{\partial\boldsymbol{\beta}}\right],$$
$$\frac{\partial \log \tilde{L}}{\partial\theta} = \sum_{i=1}^{K}\sum_{j=1}^{n_i}\left[y_{ij} - \frac{A_{ij}}{J_i}\right]\left[y_{i,j-1} + \frac{\partial\hat{\psi}(z_{ij},\boldsymbol{\alpha})}{\partial\theta}\right], \;\; \textit{and}$$
$$\frac{\partial \log \tilde{L}}{\partial\sigma^2_\tau} = \sum_{i=1}^{K}\frac{M_i}{J_i}, \tag{A.14}$$

respectively, where

$$A_{ij} = \int_{-\infty}^{\infty} \exp(\sigma_\tau\tau_i s_i)\,\Delta_i\,\tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\phi(\tau_i)d\tau_i\,, \;\; \textit{and}$$

$$M_i = \int_{-\infty}^{\infty} \exp(\sigma_\tau \tau_i s_i) \Delta_i \sum_{j=1}^{n_i} \left(\frac{\tau_i}{2\sigma_\tau} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \sigma_\tau^2} \right) \left(y_{ij} - \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \right) \phi(\tau_i) d\tau_i \,. \tag{A.15}$$

Proof. Proof is straightforward and omitted. □

Lemma A.6. *The components of $\partial^2 \log \tilde{L} / \partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top$ for (5.27) are given by*

$$\begin{aligned}
\frac{\partial^2 \log \tilde{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \sum_{i=1}^{K} \left\{ \frac{-1}{J_i^2} \left(J_i \mathbf{A}_{i\beta} + \mathbf{J}_{i\beta} \mathbf{J}_{i\beta}^\top \right) + \sum_{j=1}^{n_i} \left(y_{ij} - \frac{A_{ij}}{J_i} \right) \frac{\partial^2 \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right\}, \\
\frac{\partial^2 \log \tilde{L}}{\partial \theta^2} &= \sum_{i=1}^{K} \left\{ \frac{-1}{J_i^2} \left(J_i A_{i\theta} + J_{i\theta}^2 \right) + \sum_{j=1}^{n_i} \left(y_{ij} - \frac{A_{ij}}{J_i} \right) \frac{\partial^2 \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \theta^2} \right\}, \ \text{and} \\
\frac{\partial^2 \log \tilde{L}}{\partial {\sigma_\tau^2}^2} &= \sum_{i=1}^{K} \frac{1}{J_i^2} \left(J_i M_{i\sigma_\tau^2} - M_i J_{i\sigma_\tau^2} \right),
\end{aligned} \tag{A.16}$$

where

$$\begin{aligned}
\mathbf{J}_{i\beta} &= \sum_{j=1}^{n_i} A_{ij} \left[\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right], \\
\mathbf{A}_{i\beta} &= \int_{-\infty}^{\infty} \exp(\sigma_\tau \tau_i s_i) \Delta_i \left\{ \sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \left(1 - \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \right) \right. \\
&\quad \left[\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right] \left[\mathbf{x}_{ij}^\top + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}^\top} \right] - \left(\sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \left[\mathbf{x}_{ij} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}} \right] \right) \cdot \\
&\quad \left. \left(\sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \left[\mathbf{x}_{ij}^\top + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \boldsymbol{\beta}^\top} \right] \right) \right\} \phi(\tau_i) d\tau_i \,, \\
J_{i\theta} &= \sum_{j=1}^{n_i} A_{ij} \left[y_{i,j-1} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \theta} \right], \\
A_{i\theta} &= \int_{-\infty}^{\infty} \exp(\sigma_\tau \tau_i s_i) \Delta_i \left\{ \sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \left(1 - \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \right) \cdot \right. \\
&\quad \left. \left[y_{i,j-1} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \theta} \right]^2 - \left(\sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i) \left[y_{i,j-1} + \frac{\partial \hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial \theta} \right] \right)^2 \right\} \phi(\tau_i) d\tau_i \,,
\end{aligned}$$

$$J_{i\sigma_\tau^2} = \int_{-\infty}^{\infty} \exp(\sigma_\tau \tau_i s_i)\Delta_i \left[\frac{\tau_i}{2\sigma_\tau}s_i - \sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\left(\frac{\tau_i}{2\sigma_\tau} + \frac{\partial\hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial\sigma_\tau^2}\right)\right]\phi(\tau_i)d\tau_i\,,$$

$$\begin{aligned}
M_{i\sigma_\tau^2} = \int_{-\infty}^{\infty} \exp(\sigma_\tau \tau_i s_i)\Delta_i &\left\{\left[\sum_{j=1}^{n_i}\left(y_{ij} - \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\right)\left(\frac{\tau_i}{2\sigma_\tau} + \frac{\partial\hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial\sigma_\tau^2}\right)\right]\cdot\right.\\
&\left[\frac{\tau_i}{2\sigma_\tau}s_i - \sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\left(\frac{\tau_i}{2\sigma_\tau} + \frac{\partial\hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial\sigma_\tau^2}\right)\right]\\
&- \sum_{j=1}^{n_i} \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\left(1 - \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\right)\left(\frac{\tau_i}{2\sigma_\tau} + \frac{\partial\hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial\sigma_\tau^2}\right)^2\\
&\left.+ \sum_{j=1}^{n_i}\left(y_{ij} - \tilde{p}^*_{ijy_{i,j-1}}(\boldsymbol{\alpha}, \psi(z_{ij}), \tau_i)\right)\left(\frac{\partial^2\hat{\psi}(z_{ij}, \boldsymbol{\alpha})}{\partial{\sigma_\tau^2}^2} - \frac{\tau_i}{4\sigma_\tau^3}\right)\right\}\phi(\tau_i)d\tau_i\,. \qquad \text{(A.17)}
\end{aligned}$$

Proof. Proof is straightforward and omitted. □

Bibliography

Ahmad, I. and Lin, P.-E. (1976). A nonparametric estimation of the entropy for absolutely continuous distributions (corresp.), *IEEE Transactions on Information Theory* **22**(3): 372–375.

Al-Osh, M. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (inar (1)) process, *Journal of Time Series Analysis* **8**(3): 261–275.

Altman, N. S. (1990). Kernel smoothing of data with correlated errors, *Journal of the American Statistical Association* **85**(411): 749–759.

Andrews, D. W. (1995). Nonparametric kernel estimation for semiparametric models, *Econometric Theory* **11**(03): 560–586.

Bierens, H. J. (1987). Kernel estimators of regression functions, *Advances in econometrics: Fifth world congress*, Vol. 1, pp. 99–144.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (2007). *Discrete Multivariate Analysis: Theory and Practice*, Springer.

Blundell, R., Griffith, R. and Windmeijer, F. (1995). *Individual effects and dynamics in count data models*, Univ. College London, Department of Economics.

Breiman, L. (1992). Probability, 1968, *Reprint ed., SIAM, Philadelphia* .

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American statistical Association* **88**(421): 9–25.

Carota, C. and Parmigiani, G. (2002). Semiparametric regression for count data, *Biometrika* pp. 265–281.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American statistical association* **74**(368): 829–836.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimationg the correct degree of smoothing by the method of generalized cross-validation, *Numer. Math* **31**: 377–403.

Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures, *Biometrika* **82**(2): 407–410.

Diggle, P., Liang, K.-Y. and Zeger, S. L. (1994). Longitudinal data analysis, *New York: Oxford University Press* **5**: 13.

Eubank, R. L. (1988). Spline smoothing and nonparametric regression.

Fan, J. (1992). Design-adaptive nonparametric regression, *Journal of the American statistical Association* **87**(420): 998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies, *The Annals of Statistics* pp. 196–216.

Fan, J., Huang, T. and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function, *Journal of the American Statistical Association* **102**(478): 632–641.

Fan, J. and Wu, Y. (2008). Semiparametric estimation of covariance matrixes for longitudinal data, *Journal of the American Statistical Association* **103**(484): 1520–1533.

Farrell, P. J. and Sutradhar, B. C. (2006). A non-linear conditional probability model for generating correlated binary data, *Statistics & probability letters* **76**(4): 353–361.

Greblicki, W. and Krzyzak, A. (1980). Asymptotic properties of kernel estimates of a regression function, *Journal of Statistical Planning and Inference* **4**(1): 81–90.

Hastie, T. J. and Tibshirani, R. J. (1990). Generalized additive models, volume 43 of monographs on statistics and applied probability.

Hausman, J. A., Hall, B. H. and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r&d relationship.

Heckman, J. J. (1979). *Statistical models for discrete panel data*, Department of Economics and Graduate School of Business, University of Chicago.

Honoré, B. E. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables, *Econometrica* **68**(4): 839–874.

Horowitz, J. L. (2009). *Semiparametric and nonparametric methods in econometrics*, Vol. 12, Springer.

Jiang, J. (1998). Consistent estimators in generalized linear mixed models, *Journal of the American Statistical Association* **93**(442): 720–729.

Jowaheer, V. and Sutradhar, B. C. (2002). Analysing longitudinal count data with overdispersion, *Biometrika* **89**(2): 389–399.

Kohn, R., Ansley, C. F. and Tharm, D. (1991). The performance of cross-validation and maximum likelihood estimators of spline smoothing parameters, *Journal of the american statistical association* **86**(416): 1042–1050.

Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation, *Biometrika* **98**(2): 355.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models, *Biometrika* **73**: 13–22.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations, *Journal of the American Statistical Association* **96**(455): 1045–1056.

Lin, X. and Carroll, R. J. (2006). Semiparametric estimation in general repeated measures problems, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**(1): 69–88.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data, *Econometrica: Journal of the Econometric Society* pp. 357–362.

McCullagh, P. (1983). Quasi-likelihood functions, *The Annals of Statistics* pp. 59–67.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*, Chapman and Hall, second edn, CRC press.

McKenzie, E. (1988). Some arma models for dependent sequences of poisson counts, *Advances in Applied Probability* **20**(04): 822–835.

Montalvo, J. G. (1997). Gmm estimation of count-panel-data models with fixed effects and predetermined instruments, *Journal of Business & Economic Statistics* **15**(1): 82–89.

Moyeed, R. and Diggle, P. J. (1994). Rates of convergence in semi-parametric modelling of longitudinal data, *Australian & New Zealand Journal of Statistics* **36**(1): 75–93.

Muller, H.-G. (1988). *Non-parametric Analysis of Longitudinal Data*, Springer-Verlag.

Nadaraya, E. A. (1964). On estimating regression, *Theory of Probability & Its Applications* **9**(1): 141–142.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society: Series A* **135**: 370–384.

Pagan, A. and Ullah, A. (1999). *Nonparametric econometrics*, Cambridge university press.

Picard, R. R. and Cook, R. D. (1984). Cross-validation of regression models, *Journal of the American Statistical Association* **79**(387): 575–583.

Powell, J. L. and Stoker, T. M. (1996). Optimal bandwidth choice for density-weighted averages, *Journal of Econometrics* **75**(2): 291–316.

Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses, *Biometrics* pp. 825–839.

Rao, R. P., Sutradhar, B. and Pandit, V. (2012). Gmm versus gql inferences in semiparametric linear dynamic mixed models, *Brazilian Journal of Probability and Statistics* pp. 167–177.

Royall, R. M. (1966). *A class of non-parametric estimates of a smooth regression function.*, PhD thesis, Dept. of Statistics, Stanford University.

Ruppert, D. (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation, *Journal of the American Statistical Association* **92**(439): 1049–1062.

Schall, R. (1991). Estimation in generalized linear models with random effects, *Biometrika* pp. 719–727.

Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models, *Journal of the American statistical Association* **89**(426): 501–511.

Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models, *The Annals of statistics* pp. 1768–1802.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Vol. 26, CRC press.

Sneddon, G. and Sutradhar, B. C. (2004). On semiparametric familial–longitudinal models, *Statistics & probability letters* **69**(3): 369–379.

Speckman, P. (1988). Kernel smoothing in partial linear models, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 413–436.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models, *Journal of the American Statistical Association* **84**(405): 276–283.

Stone, C. J. (1977). Consistent nonparametric regression, *The annals of statistics* pp. 595–620.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators, *The annals of Statistics* pp. 1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *The annals of statistics* pp. 1040–1053.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions, *Journal of the royal statistical society. Series B (Methodological)* pp. 111–147.

Sutradhar, B. C. (2003). An overview on regression models for discrete longitudinal responses, *Statistical Science* **18**(3): 377–393.

Sutradhar, B. C. (2004). On exact quasilikelihood inference in generalized linear mixed models, *Sankhyā: The Indian Journal of Statistics* pp. 263–291.

Sutradhar, B. C. (2010). Inferences in generalized linear longitudinal mixed models, *Canadian Journal of Statistics* **38**(2): 174–196.

Sutradhar, B. C. (2011). *Dynamic Mixed Models for Familial Longitudinal Data*, Springer Series in Statistics, Springer.

Sutradhar, B. C. and Bari, W. (2007). On generalized quasilikelihood inference in longitudinal mixed model for count data, *Sankhyā: The Indian Journal of Statistics* pp. 671–699.

Sutradhar, B. C., Bari, W. and Das, K. (2010). On probit versus logit dynamic mixed models for binary panel data, *Journal of Statistical Computation and Simulation* **80**(4): 421–441.

Sutradhar, B. C. and Das, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data, *Biometrika* **86**: 459–465.

Sutradhar, B. C. and Farrell, P. J. (2007). On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data, *Sankhyā: The Indian Journal of Statistics* pp. 448–467.

Sutradhar, B. C., Jowaheer, V. and Rao, R. P. (2014). Remarks on asymptotic efficient estimation for regression effects in stationary and nonstationary models for panel count data, *Brazilian Journal of Probability and Statistics* **28**(2): 241–254.

Sutradhar, B. C., Rao, R. P. and Pandit, V. (2008). Generalized method of moments versus generalized quasilikelihood inferences in binary panel data models, *Sankhyā: The Indian Journal of Statistics, Series B (2008-)* pp. 34–62.

Sutradhar, B. C., Warriyar, K. and Zheng, N. (2016). Inference in semi-parametric dynamic models for repeated count data, *Australian & New Zealand Journal of Statistics* **58**(3): 397–434.

Ten Have, T. R. and Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary response data, *Biometrics* **55**(1): 85–93.

Wahba, G. (1990). *Spline models for observational data*, SIAM.

Warriyar, V. and Sutradhar, B. C. (2014). Estimation with improved efficiency in semi-parametric linear longitudinal models, *Brazilian Journal of Probability and Statistics* **28**(4): 561–586.

Watson, G. S. (1964). Smooth regression analysis, *Sankhyā: The Indian Journal of Statistics, Series A* pp. 359–372.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the gauss-newton method, *Biometrika* **61**(3): 439–447.

Whittaker, E. T. (1922). On a new method of graduation, *Proceedings of the Edinburgh Mathematical Society* **41**: 63–75.

Winkelmann, R. (2008). *Econometric analysis of count data*, Springer Science & Business Media.

Wooldridge, J. M. (1999). Distribution-free estimation of some nonlinear panel data models, *Journal of Econometrics* **90**(1): 77–97.

You, J. and Chen, G. (2007). Semiparametric generalized least squares estimation in partially linear regression models with correlated errors, *Journal of Statistical Planning and Inference* **137**(1): 117–132.

Zeger, S. L. (1988). A regression model for time series of counts, *Biometrika* **75**(4): 621–629.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to cd4 cell numbers in hiv seroconverters, *Biometrics* pp. 689–699.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a gibbs sampling approach, *Journal of the American statistical association* **86**(413): 79–86.

Zeger, S. L., Liang, K.-Y. and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach, *Biometrics* pp. 1049–1060.

Zeger, S. L., Liang, K.-Y. and Self, S. G. (1985). The analysis of binary longitudinal data with time-independent covariates, *Biometrika* pp. 31–38.

Zhao, L. P. and Prentice, R. L. (1990). Correlated binary regression using a quadratic exponential model, *Biometrika* **77**(3): 642–648.

Zheng, N. and Sutradhar, B. C. (2016). Inferences in semi-parametric dynamic mixed models for longitudinal count data, *Annals of the Institute of Statistical Mathematics* pp. 1–33.